

What's wrong with UNESCO

The new director-general needs to buck all expectations and transform the agency.

As *Nature* went to press, Irina Gueorguieva Bokova, a Bulgarian diplomat, and Farouk Hosny, Egypt's minister of culture, faced off in a final electoral round to become director-general of the United Nations Educational, Scientific and Cultural Organization (UNESCO). The vote, by the agency's executive board, is subject to confirmation in October by its general conference.

The election highlights UNESCO's faults. The agency publishes a list of the candidates' names, but nothing on their qualifications or vision for the agency. In truth, nominations and the secret voting are largely down to horse-trading among member states.

Such anachronistic processes need to change, but so does much else at this suffocatingly bureaucratic organization. An independent review of UNESCO's science portfolio in 2007 reached the damning conclusion that "UNESCO has over time lost its leadership credibility as an international spokesman for science, and its programmes are now seen by the scientific community as fragmented, over-ambitious, unfocused and lacking a clear vision and scientific strategy".

That should have been a wake-up call. But the panel's suite of recommendations, including prioritizing and refocusing the agency's science activities, and establishing an advisory committee of outside scientists reporting to the director general (see <http://tinyurl.com/unesco-science>), met with a defensive response. The agency rejected the idea of an expert committee, while implementation of the panel's other recommendations (<http://tinyurl.com/unesco-genconf>) does not go far enough. Business as usual continues.

UNESCO's science section has an annual regular budget of around US\$55 million and 160 staff. It needs to focus its efforts on a few areas where it might make an impact. With science-based issues now omnipresent throughout the United Nations (UN) and many other international agencies, UNESCO must slash everything superfluous, such as its puny programme on renewable energy.

UNESCO has strengths to build on. Its water programme is large and flabby, but outside scientists rate some parts of it highly, such as the International Hydrological Programme on water research and

management. These should be reinforced and integrated with the activities of the two dozen other UN agencies that work on water. Too often, UNESCO's science programmes are isolated from related work elsewhere, and even from its own social science, education, and culture arms. The agency also has an important resource in its network of national offices, and UNESCO-branded field centres such as biosphere reserves and science labs, where many potential synergies, for example for ecological monitoring, are under-exploited.

Yet despite its shortcomings, UNESCO is uniquely placed in being the only UN agency with an explicit mandate to promote science. And its intergovernmental status, although often a handicap, potentially gives it the power to convene the world's best expertise to take forward important agendas.

UNESCO has made a start along those lines. Its advice to Nigeria on building a science system is credited as a factor in the Nigerian government's \$5 billion commitment to science in 2006. UNESCO has the potential to become a leader in such areas, providing policy analysis and benchmarking for less scientifically advanced countries. This seems a better road to promoting infrastructure than its current smattering of tiny grants in its International Basic Sciences Programme. UNESCO should give up the hopeless notion that it can be a research funder, and focus on policy and leverage.

The outgoing director general Koïchiro Matsuura, a Japanese diplomat, has reformed UNESCO's finances and recruitment practices. But he brought little vision or change to the science programme. His successor should take the 2007 review as the starting point for a root-and-branch review of the science programme, persuade the member states to weed out all activities that have little or no impact and create a culture of performance, transparency and evaluation. An upcoming wave of retirement at the agency provides an opportunity to bring in fresh blood.

The history and culture of UNESCO do not bode well for serious change. But business as usual is not an option if UNESCO is to have a scientific *raison d'être*. ■

Earth's boundaries?

An attempt to quantify the limits of humanity's load on our planet opens an important debate.

In this issue of *Nature*, a group of renowned Earth-system and environmental scientists led by Johan Rockström of the Stockholm Resilience Centre sets out to define boundaries for the biophysical processes that determine the Earth's capacity for self-regulation (see page 472). The framework presented is an attempt to look holistically at how humanity is stressing the entire Earth system. Provocatively, they go beyond the conceptual to suggest numerical boundaries for

seven parameters: climate change, ozone depletion, ocean acidification, biodiversity, freshwater use, the global nitrogen and phosphorus cycles, and change in land use. The authors argue that we must stay within all of these boundaries in order to avoid catastrophic environmental change.

The boundaries are based on existing data. For some processes, such as anthropogenic climate change and human modification of the nitrogen cycle, we may already have crossed the line, and need to back-pedal quickly. For others, such as ocean acidification, we are rapidly approaching a threshold beyond which there may be abrupt and nonlinear changes.

The exercise requires many qualifications. For the most part, the exact values chosen as boundaries by Rockström and his colleagues

are arbitrary. So too, in some cases, are the indicators of change. There is, as yet, little scientific evidence to suggest that stabilizing long-term concentrations of carbon dioxide at 350 parts per million is the right target for avoiding dangerous interference with the climate system. Focusing on long-term atmospheric concentrations of the greenhouse gas is perhaps an unnecessary distraction from the much more immediate target of keeping warming to within 2 °C above pre-industrial levels. Nor is there a consensus on the need to cap species extinctions at ten times the background rate, as is being advised.

Furthermore, boundaries don't always apply globally, even for processes that regulate the entire planet. Local circumstances can ultimately determine how soon water shortages or biodiversity loss reach a critical threshold.

Assigning 'acceptable' limits to processes that ultimately determine our own survival is risky in other ways, too. After all, some of the suggested limits may be easier to balance with ethical and economic issues than others. Human interference in the nitrogen cycle may well have damaging long-term consequences, but the production of

nitrogen for agriculture has also fed large parts of humanity.

But even if the science is preliminary, this is a creditable attempt to quantify the limitations of our existence on Earth, and provides a good basis for discussion and future refinement. To facilitate that discussion, *Nature* is simultaneously publishing seven commentaries from leading experts that can be freely accessed at *Nature Reports Climate Change* (see <http://tinyurl.com/planetboundaries>).

Defining the limits to our growth and existence on this planet is not only a grand intellectual challenge, it is also a potential source of badly needed information for policy-makers. Such numerical values, however, should not be seen as targets. If the history of environmental negotiations has taught us anything, it is that targets are there to be broken. Setting limits that are well within the bounds of linear behaviour might therefore be a wiser, if somewhat less dramatic, approach. That would still give policy-makers a clear indication of the magnitude and direction of change, without risking the possibility that boundaries will be used to justify prolonged degradation of the environment up to the point of no return. ■

Biobanks need pharma

Which is why Europe's citizens need reassurance that their donations will be in the public interest.

Medical geneticist Thomas Meitinger remembers when biobanking was a simple craft. As a postdoc thirty years ago, he travelled from Oxford to Yugoslavia to track down a family afflicted with a rare disease causing blindness. The family listened enthusiastically as he explained his research over a fish dinner. He returned with blood samples and over the next decade used them to identify the single gene defect that caused the condition.

Biobanking — collecting tissue or body fluids alongside medical information — is now a large-scale affair. Genomics allows geneticists to track down not just the single genes that convey a strong risk of disease, but also the many low-risk genes associated with the diseases that kill most of us, such as cancer, diabetes and cardiovascular disease. But these very weak gene signals can be picked up only in studies of large populations of up to hundreds of thousands of people.

Europe leads the world in biobanking. It has more than 400 biobanks, some involving hundreds of thousands of diseased and healthy individuals. It is now seeking to make the most of that resource: the European Commission is funding a preparatory study aimed at linking the biobanks into one distributed infrastructure. Now Meitinger, who currently works at the Institute of Human Genetics in Neuherberg, Germany, and the rest of the scientific consortium driving the effort, called the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI), must find stable funding for the project and arrange access for the scientific community.

That's a lot of tough challenges at a time when the general public is sensitive to any issue involving genes and biological material. Key concerns in biobanking are those of anonymity and whether true informed consent can be given by individual donors now too numerous to be educated over dinner.

Another, potentially incendiary, issue is whether the pharmaceutical industry should have the same access rights to biobanks as academic researchers. Europe's citizens could easily turn against biobanking if they start to feel exploited for financial gain. The BBMRI must accommodate industry while avoiding such a backlash.

Biobank resources may be fundamental to understanding the molecular bases of common complex diseases, but it is the pharmaceutical industry that will develop the treatments for such ailments. Companies generate their own biobanks, but these cannot reach the scale necessary to move forward. Industry wants access to large public biobanks, and the BBMRI recognizes its obligation to facilitate new medicines. The consortium hopes that relentless outreach and appropriate control of banked materials will achieve this without antagonizing the public.

At the consortium's first stakeholder meeting last week, patient groups declared that they don't care who gets hold of their diseased tissue "so long as it is out of our bodies and being used to do clinical good". But the large majority of healthy donors will need more persuading that profit-making industry should get access to their voluntary tissue donations.

The concept of expert centres, unveiled by the BBMRI at the meeting, should help. These would do all the molecular analyses on material requested for an approved study and provide data only to clients. Donors' material would not move out of the biobanking infrastructure, and data would be stored for re-use in other studies, so industry could not gain exclusive rights.

Industry must also be prepared to give something back, in the form of access to its own biobanks and their richly financed expertise. Research departments across all companies believe that biobanks and the molecular information generated from them are outside the competitive realm, but their managers tend to be wedded to secrecy. So managers must be persuaded to follow their researchers' instincts, before the public gets the idea that industry is there only to exploit, gets deterred from donating, and the whole enterprise becomes tainted with distrust. ■

RESEARCH HIGHLIGHTS

Getting thinner faster

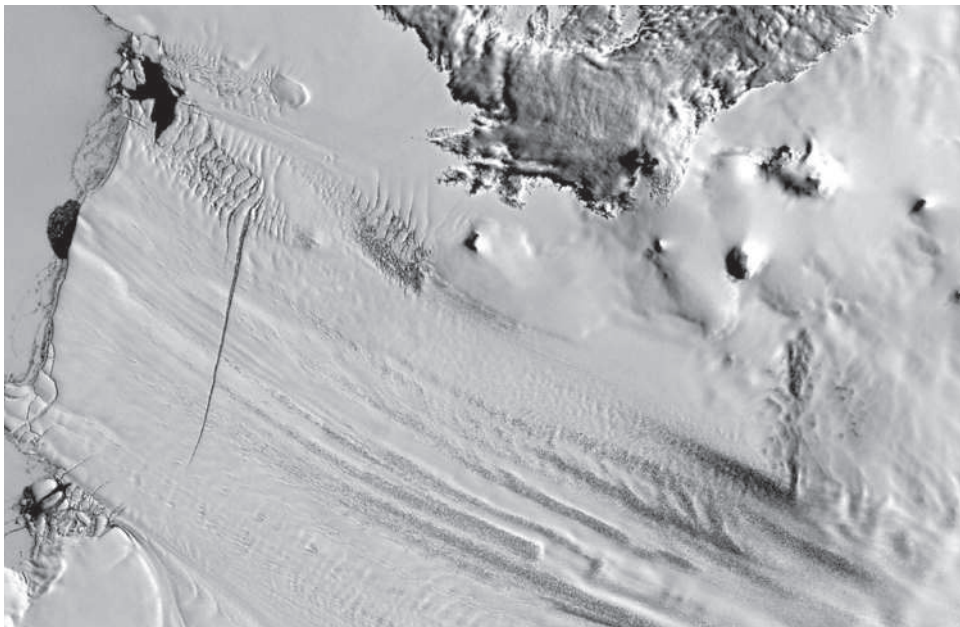
Geophys. Res. Lett. **36**, L17501

doi:10.1029/2009LL039126 (2009)

The rate of thinning of the Pine Island Glacier (pictured) — the largest stream of fast-moving ice on the West Antarctic Ice Sheet — quadrupled from 1995 to 2006. If the current rate of acceleration continues, the main trunk of the glacier could be afloat within 100 years.

Using satellite radar altimetry to examine the ice thickness, Duncan Wingham of University College London and his colleagues found that the thinning process has rapidly propagated inland since 1995, with the frozen tributaries that flow into the central trunk of the ice stream now losing mass as well.

Scientists think that the imbalance has been triggered by warm waters at the glacier's ocean terminus.



NASA

IMMUNOLOGY

Natural born killers

Nature Immunology doi:10.1038/ni.1787 (2009)

A master gene turns blood stem cells into the 'natural killer' immune cells that fight viral infections and help stave off cancer.

When Hugh Brady at Imperial College London and his colleagues created knockout mice lacking the gene for a particular gene-regulatory protein, E4bp4, the mice made the other immune system cells — B cells and T cells — but no killer cells. Adding back the gene for E4bp4 to blood stem cells from E4bp4-deficient mice enabled these cells to produce natural killer cells again.

The mouse strain lacking E4bp4 will help in investigating the role of natural killer cells in disease prevention and in finding drugs that boost their production, the authors say.

SYSTEMS BIOLOGY

Metabolic map

Science **325**, 1544–1549 (2009)

Researchers have reconstructed the metabolic pathways of the heat-loving bacterium *Thermotoga maritima*, which is found in underwater volcanic vents. The work reveals a network of 478 proteins that contain only 182 basic shapes or 'folds'.

Adam Godzik of the Burnham Institute for Medical Research in La Jolla, California, and his colleagues combined structural-genomics and systems-biology approaches to show that most of these folds are found in proteins that are encoded by the most

essential core genes, indicating that the folds perform specific chemical functions that are needed for proper metabolism.

EVOLUTION

Armed and dangerous

Proc. R. Soc. Lond. B doi:10.1098/rspb.2009.1256 (2009)

What good is a horn on a female yak (pictured, below)? Naturalists have long known that competition for mates drove the evolution of horns among male ungulates, but the presence of prongs on some females has defied coherent explanation ever since Charles Darwin puzzled over them.

Now, Theodore Stankowich of the University of Massachusetts and Tim Caro of

the University of California at Davis propose that, in the case of large animals living in open territory, the 'weaponry' on female bovids arose mainly for defence against predators. In smaller bovids, they say, female horns evolved for territorial battles among females of the same species. These two hypotheses agree with a phylogenetic reconstruction, and explain the presence or absence of horns in 80 of the 82 bovid species compared.

BEHAVIOURAL SCIENCE

Jelly shots and jackpots

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0906629106 (2009)

Rats fed tasty 'jelly shots' containing alcohol during adolescence became bigger risk-takers than teetotaler rats when presented with a lever game designed by Ilene Bernstein and her colleagues at the University of Washington in Seattle. When the adult rats were faced with a choice between pressing a lever for a guaranteed two sugar pills or a lever that could give them either nothing or four sugar pills, the individuals exposed to alcohol in adolescence tended to gamble more often. This effect on behaviour could still be seen three months after the alcohol was discontinued.

The researchers believe that their results indicate that the risk-taking behaviour is caused by the alcohol. Previous studies were not able to rule out the possibility that alcohol use and risk-taking behaviour are both caused by the same underlying propensity.



C. BOISVIEUX/CORBIS

NEUROBIOLOGY

Shocktopus

Curr. Biol. doi:10.1016/j.cub.2009.07.067 (2009)

Octopuses have the peculiar challenge of controlling eight appendages that can assume an almost limitless number of positions. They seem to accomplish this by consolidating the control of complex, coordinated movements into specific areas of their nervous system — rather than having a one-to-one relationship between a body part and a specific brain area.

Letizia Zullo, now at the Italian Institute of Technology in Genova, and her colleagues placed electrodes in 35 parts of the animal's nervous system. Low-voltage stimulation of different areas evoked simple responses, such as a change in skin colour or small eyelid movements. Higher voltages elicited more complex responses, such as inking and jet-propelled swimming.

GENE THERAPY

Panning for phage

Nature Medicine doi:10.1038/nm.2025 (2009)

Viruses can be modified to deliver therapeutic genes directly to the blood vessels feeding diseased brains, according to research by Beverly Davidson and her colleagues at the University of Iowa in Iowa City. To identify address labels for the diseased blood vessels, the team used the technique of 'phage panning'. Millions of particles of a bacterial virus — phage — displaying different mouse proteins on their surface were injected into the blood of mice with a lysosomal storage disease. Phages that stuck to diseased-brain blood vessels were isolated and the protein sequences responsible for binding were identified.

When these sequences were engineered into the outer shell of an adeno-associated virus, AAV2, they enabled AAV2 to target brain

blood-vessel cells in the diseased mice and deliver a gene coding for an enzyme that helps to reverse the effects of the storage defect.

EVOLUTIONARY BIOLOGY

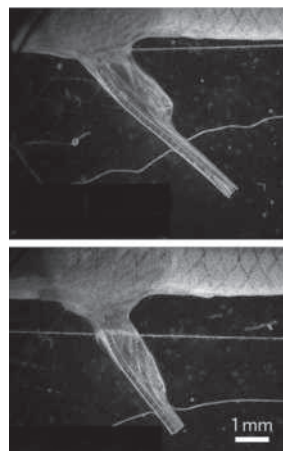
Well endowed

Biol. Lett. doi:10.1098/rsbl.2009.0637 (2009)

The wide variation in male genitalia size in animals is thought to have evolved mostly in response to selection pressures that come into play during or after copulation and increase the male's share of paternity.

But it seems that females of the mosquito-fish *Gambusia holbrooki* choose mates before copulation on the basis of the size of their genitalia — and for them, bigger is better.

Andrew Kahn and his team at the Australian National University in Canberra tested female preference for males that had had their genitalia considerably reduced in size by surgery (pictured, bottom) compared with those with only a minor reduction (pictured, top). They found that females spent, on average, around one-and-a-half times longer associating with the better-endowed males.



ANALYTICAL CHEMISTRY

Evaporating flesh

Angew. Chem. Int. Ed. doi:10.1002/anie.200902546 (2009)

Cancer surgeons cutting out tumours in the operating theatre can find it hard to spot the difference between malignant and healthy tissue. But Hungarian scientists say it should be possible to identify tissue types in real time

during surgery, using a mass spectrometer to analyse the ions sprayed up by a standard surgical electrode as its electric current evaporates living flesh.

Zoltán Takáts at Semmelweis University, Budapest, and his colleagues discovered that different tissues yield characteristic sets of gaseous molecular ions as they are being dissected — and that the ions can be pumped to a mass spectrometer and analysed in less than a second to distinguish various grades of cancerous tissue.

AQUATIC TOXICOLOGY

Mixed-up fish

Aquatic Toxicol. doi:10.1016/j.aquatox.2009.08.001 (2009)

Pollutants such as effluent wastewater, pesticides and pharmaceuticals in rivers are suspected to cause the intersex condition in fish, in which reproductive tissues are mismatched: most often, female germ cells are

present in the testes.

To assess the incidence of the condition in river fish throughout the United States, Jo Ellen Hinck, a US Geological Survey scientist in Columbia, Missouri, and her team sampled 111 sites spread across the country. Intersex fish were found in 31% of the sites, with the greatest prevalence in the south-east. Small and largemouth bass were the species most commonly affected.

Correction

The Research Highlight 'Aluminium arches' (*Nature* 461, 318; 2009) incorrectly referred to an aluminium powder used in the experiment. In fact, an aluminium oxide powder, alumina, was used.

JOURNAL CLUB

Mikiko C. Siomi
Keio University School of
Medicine, Tokyo, Japan

A biologist praises a mouse model of autism inheritance.

Autism, a neurodevelopmental disorder that affects people's social abilities, has both genetic and non-genetic causes. Chromosomal abnormalities account for 10–20% of autism cases, with duplication of a long stretch of chromosome 15 being the most common. I was excited to read

that a mouse model with a similar chromosomal duplication has been generated (J. Nakatani *et al.* *Cell* 137, 1235–1246; 2009). These mice exhibit the inflexible behaviour, social abnormalities and increased anxiety often observed in people with autism. However, whereas the engineered mice inherit the duplication from their fathers, human autism cases caused by such a duplication are usually inherited maternally. Further genomic analysis in the mice should find the reason for this discrepancy.

This model deserves special attention as the chromosomal

duplication is stably maintained between generations. Also, genes in the duplicated regions seem to work; that is, the expression levels of genes — including *HBI152*, which affects the function of serotonin, a molecule that has cognitive roles in mood, memory and learning — are higher in the mice, as would be expected with a gene duplication.

Accumulated evidence shows that variations in gene-copy number, such as the chromosomal duplication in this model, are associated with susceptibility to various human diseases; cancer cells, for example, tend to have

high gene-copy numbers. Thus, this mammalian model may help us to understand the molecular basis of autism and to investigate the contribution of gene duplication in other genetic diseases. This should encourage many researchers to produce other model systems for copy-number variation using similar techniques; systems that may clarify the contribution of chromosomal duplication, or even the lack of it, in common diseases such as diabetes.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

NEWS BRIEFING

● POLICY

Science advice: Europe will have a **chief scientist**, European Commission President José Manuel Barroso has promised. He made the pledge in a 15 September speech at the opening of the newly elected European Parliament, which also voted to extend his presidency for a second term. Barroso added that he would review the way the commission uses scientific advice — a move welcomed by researchers. For more, see <http://tinyurl.com/chiefscientist>

Refrigerant curb:

Hydrofluorocarbons (HFCs) inched closer to regulation as greenhouse gases under the **Montreal Protocol**, after the **United States, Canada and Mexico** joined a growing movement to amend the treaty. HFCs, which are used as refrigerants, were introduced to replace chemicals that destroy stratospheric ozone. Their regulation falls under the United Nations climate framework, but many say they could be phased out faster and more cheaply using the Montreal ozone treaty. The issue could be decided as early as November, when Montreal Protocol delegates convene in Egypt for their annual meeting. For more, see <http://tinyurl.com/hfccurb>

Vaccine sharing: Nine countries have pledged to share up to 10% of their pandemic H1N1 flu vaccines with **developing nations**. Australia, Brazil, France, Italy, New Zealand, Norway, Switzerland, the United States and Britain all said they would make vaccines available through the World Health Organization, which has been asking for such donations.

Biology initiative: The US National Research Council called last week for a new inter-agency, multidisciplinary life-sciences programme that would tackle problems in the fields of food, environment, energy and health.

ARCTIC ICE

Arctic sea ice has reached its third-lowest level since satellite radar measurements began in 1979. Scientists with the University of Colorado's National Snow and Ice Data Center in Boulder announced on 17 September that the sea-ice extent of 5.1 million square kilometres observed on 12 September was the low point for the year. This year's minimum is higher than in both 2007 and 2008, but does not indicate any reversal in a marked 30-year decline in summertime ice extent, scientists said. For more, see <http://tinyurl.com/seaicelow>



S. FLOOD/GETTY

The 10-year “**new biology**” **initiative** should be supported by new dedicated funding, not from existing agency research budgets, the report from the Board on Life Sciences urged. Report co-author Keith Yamamoto of the University of California, San Francisco, estimates that at least US\$2 billion per year should be earmarked for the scheme.

Research assessment: Details of a new way to assess the quality of research in UK universities were published for consultation on 23 September. The proposed **Research Excellence Framework** will inform decisions on how to carve up around £1.76 billion (US\$2.9 billion) per year between institutions, and will use bibliometrics — such as citation counts — to aid peer-reviewed judgements of work quality. It is due to come into force in 2013, replacing the Research Assessment Exercise. For more, see <http://tinyurl.com/resframe>.

Climate clamour: Ahead of United Nations climate meetings in preparation for the Copenhagen summit, a flurry of reports warned last week of

the impact of climate change on the developing world. Alongside calls for strong political leadership came the **World Bank 2010 Development Report**, a biennial assessment which estimated that poor nations would need more than US\$500 billion a year to adapt to climate change and develop low-carbon technology — in line with earlier UN estimates. And in a letter to the *British Medical Journal* and *The Lancet*, doctors warned of a “global health catastrophe” resulting from climate change — with effects ranging from widespread malnutrition to the spread of tropical diseases.

NUMBER CRUNCH 50%

The aviation industry pledged to halve its carbon dioxide emissions from 2005 levels by 2050 at the United Nations' climate summit in New York

● BUSINESS

Stem-cell trial: The biotherapeutics firm Neuralstem in Rockville, Maryland, has received permission from the US Food and Drug Administration to inject neural stem cells into the spinal cords of 12 patients with amyotrophic lateral sclerosis (Lou Gehrig's disease). The trials — the first stem-cell approach for the condition — will focus on safety rather than efficacy, and are expected to take place

at Emory University, Georgia, subject to approval from the university's patient-safety board. For more, see <http://tinyurl.com/Neuralstem>

RESEARCH

Resignation: The head of research at the **Swiss Federal Institute of Technology (ETH)**, Zurich, has resigned over concerns that spectroscopic data published by his research group in 1999 and 2000 had been faked. Peter Chen, a physical organic chemist, stepped down as vice-president for research after a commission that he had requested in January concluded that two papers and a doctoral thesis contained falsified data. Chen denies any involvement in handling the data and is retaining his post as a professor. For more, see <http://tinyurl.com/eth-resig>

Stem cells: The **US National Institutes of Health (NIH)** began accepting applications on 21 September to evaluate which human embryonic stem cells will be eligible for federal dollars. A panel of nine scientists, lawyers and ethicists will check that the submissions meet new requirements for informed consent from embryo donors. The working group has not yet appraised any cell lines — including the 21 lines approved under former US President George W. Bush, which must be reassessed — and will start considering particular lines after scientists submit their petitions on the NIH website.

NEWS MAKER



Arun Majumdar

The director of the environmental energy department at Lawrence Berkeley National Laboratory in California has been nominated to head the US Advanced Research Projects Agency-Energy.

Plague death: A molecular geneticist at the **University of Chicago** died on 13 September after exposure to an attenuated strain of *Yersinia pestis*, the bacterium that causes plague. Malcolm Casadaban, 60, was

working with a strain of the bacterium that has been used in vaccines in some countries and is approved by the US Centers for Disease Control and Prevention in Atlanta, Georgia, for laboratory studies. He died about 12 hours after being admitted to hospital. The bacterium was not identified as the cause of death until 18 September.

Planck on course: The Planck telescope enjoyed a successful test run, said the **European Space Agency (ESA)**, as it geared up to map the sky's cosmic microwave background. ESA reported on 17 September that preliminary results showed excellent data quality.

Stem-cell collaboration: California and Germany are collaborating on stem-cell research. The California Institute for Regenerative Medicine and the German Federal Ministry of Education and Research announced plans last week to work together on joint research projects, focusing on immunology. The agreement makes it easier for researchers in both jurisdictions to apply for joint funding. The California stem-cell agency already has international collaborations with Australia, Canada, Japan, Spain and the United Kingdom.

European appointment: Gérard de Nazelle, a former global manager for research and innovation at Shell, has been appointed the first director of the European Institute of Innovation and Technology, at its Budapest headquarters.

THE WEEK AHEAD

28 SEPTEMBER–9 OCTOBER

The United Nations Framework Convention on Climate Change holds its fourth round of international climate talks this year, this time in Bangkok.

► <http://tinyurl.com/climatetalks>

30 SEPTEMBER

A US National Academy of Sciences panel releases a report on security and screening measures for labs working with biological agents and toxins.

30 SEPTEMBER

By the end of the US government's fiscal year, the National Institutes of Health will have allocated a considerable chunk of its allotted US\$10.4 billion in stimulus funds. A rush of grant announcements is due.

AWARDS

Enrico Fermi Award: This year's winners, announced on 17 September, were **John Goodenough**, now at the University of Texas at Austin, and **Siegfried Hecker** at Stanford University, California, and previously director of the Los Alamos National Laboratory in New Mexico. They share US\$375,000. Goodenough developed cathode materials for the lithium-ion rechargeable battery; Hecker was cited for his research on plutonium metallurgy, and his leadership on global nuclear non-proliferation.

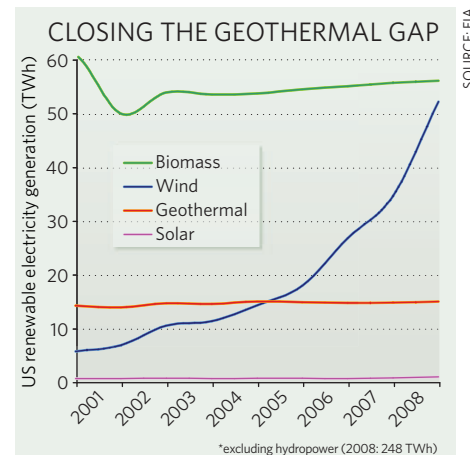
BUSINESS WATCH

The US geothermal industry is anticipating a boost in investment, as the Department of Energy prepares to announce recipients of US\$400 million in stimulus grants for research and demonstration projects in the sector.

Geothermal energy is a consistent source of clean power, but electricity investors are not used to risking the large up-front cost of exploratory drilling. The department's money will "help drill a lot of holes", says Ethan Zindler, of consultants New Energy Finance in London. Geothermal's contribution to US electricity generation (see chart) has been constant for more than a decade. But the country's 3 gigawatts of installed capacity — the largest of any nation — could double over

the next ten years if projects already planned work out, says Karl Gawell, executive director of the Geothermal Energy Association in Washington DC.

In July, Magma Energy, based in Vancouver, Canada, raised more than Can\$110 million (US\$103 million) in an initial public offering. Private geothermal developer Ram Power, in Reno, Nevada, raised Can\$180 million by private subscriptions and is expecting to merge with three Canadian geothermal firms. But in California, a US\$17-million project was halted earlier this month after operator AltaRock Energy in Sausalito encountered problems at its drilling site.



Buoy damage blurs El Niño forecasts

Missing data from the eastern Pacific Ocean may hinder predictions of this year's event.

More than half a dozen oceanographic buoys are missing in action in the eastern Pacific Ocean. The US National Oceanic and Atmospheric Administration (NOAA) has dispatched a ship to fix the malfunctioning buoys, which are part of the 55-strong Tropical Atmosphere Ocean (TAO) array that monitors the ocean and atmosphere.

But the data void has left climate modellers uncertain about the strength of this winter's El Niño — which occurs when warm waters shift east across the Pacific, bringing rain and extreme weather.

Fishermen, storms and boat collisions often knock buoys in this region out of commission. The upcoming El Niño adds “more urgency than usual”, says Arun Kumar, branch chief at NOAA's Climate Prediction Center in Camp Springs, Maryland.

Among other things, the TAO array tracks the 20°C thermocline, a boundary between warm surface and cool deep waters that ranges from about 140 metres deep in the western Pacific to 40 metres deep in the eastern Pacific. Shifts up and down in the thermocline's location can indicate the magnitude of an incipient El Niño or a La Niña, the cold phase of the climate oscillation.

Because changes to the thermocline are more pronounced in the shallower waters of the eastern Pacific, the buoys there provide “an important chunk of data from a region where the signals tend to be very large during El Niño and where the models depend on the data to forecast accurately”, says Michael McPhaden, an oceanographer at NOAA's Pacific Marine Environmental Laboratory in Seattle, Washington. “We are at a critical point in El Niño — everyone realizes we have to get those buoys up and working.”

The two lines of buoys in the fish-rich region of the eastern Pacific have a history of being tinkered with, and for each of the past four major El Niño events, there have been at least five missing. “But 2009 is exceptional,” says McPhaden.

Some of the closest buoys are about a week's sail from the coast. Fishermen often tie up to a buoy and drag it off to the side, then sweep their nets through the water to catch the



The buoy array needs regular maintenance.

fish that gather there. Past repair trips have found sliced mooring lines, fishing nets tangled on the instruments and bullet holes on the buoys themselves — and sometimes, no buoys left at all, says Shannon McArthur, the TAO project manager at the National Data Buoy Center in Mississippi.

In one instance, a buoy stopped transmitting temperature and other data, but it continued to relay its position — allowing scientists to watch as it drifted for a day, then took off at high speed to land at Puntarenas, Costa

Rica, probably taken aboard a ship. McArthur will describe this and other details of buoy vandalism at the Oceans '09 meeting in Biloxi, Mississippi, from 26 to 29 October.

Half of the 14 TAO buoys that are strung along two mooring lines, along the longitudes of 95° west and 110° west, have stopped transmitting in the past eight months. The research vessel *Wecoma* set off on 4 September from Newport, Oregon, to fix them — a job that could take a month. The cost of replacing or fixing buoys depends on their location and how much monitoring equipment they carry, says McArthur. NOAA spends an estimated US\$1 million each year to repair the array.

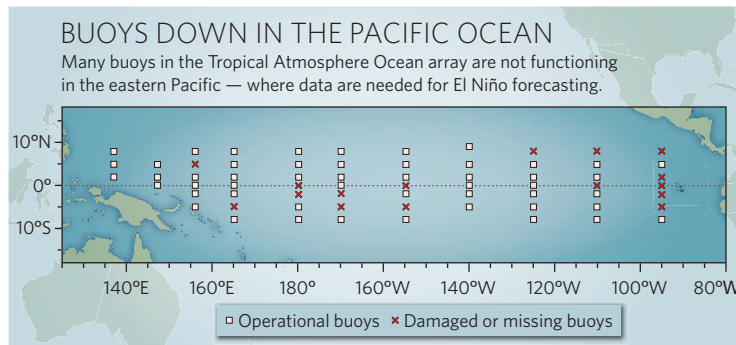
And although lines of buoys are scheduled for maintenance every six months, this year the swathe in the eastern Pacific slipped two months behind that schedule. McArthur says that repairs aren't late, but that financial and other logistical concerns can sometimes affect scheduled maintenance. A NOAA effort is under way to calculate costs for the buoy-operating community at large, he says, but it's difficult to quantify vandalism's impact on data.

Meanwhile, forecasters are scrambling to work out how the missing data will affect their El Niño predictions. In August, NOAA predicted a mild El Niño this autumn that will strengthen through the winter, but other models forecast a more extreme event.

“Because of the missing moorings, the forecasting system must now rely more heavily on the other observing systems,” says Magdalena Balmaseda, a senior scientist at the European Centre for Medium-Range Weather Forecasts in Reading, UK, who has published on the lack of redundancy in data-collecting systems in the Pacific (M. Balmaseda and D. Anderson *Geophys. Res. Lett.* **36**, L01701; 2009).

For now, forecasters are patching in data from buoys farther to the west to try to improve predictions in the eastern Pacific. Balmaseda says that the data problem can be worse when altimeter instruments aboard satellites stop working, as happened in August 2008. For now, she says, “we still have hope.”

Naomi Lubick



See News & Views, page 481, and Letters, page 511.

SPECIAL REPORT

German science looks to new political players

Coalition change could affect policies, reports **Quirin Schiermeier**.

Science budgets in Germany have flourished under the current government led by Chancellor Angela Merkel. And despite the gloomy economic climate, the 27 September parliamentary elections seem unlikely to change that.

In fact, the biggest changes to German science could come from policies advocated by the Free Democrats, who may end up as the new junior partner in an extension of Merkel's Christian Democrat government. The Free Democrats, who could replace the Social Democrats in coalition with the Christian Democrats, have a distinctly liberal approach in hot-button areas such as genetically modified crops and stem cells.

All recent polls show Merkel's planned new coalition to have an average share of the vote of around 50%. The opposing Social Democrat party, which has nominated foreign minister Frank-Walter Steinmeier as its candidate for chancellor, has dropped to a historic low of 25% in the polls, with the

Greens and the Left Party behind that.

Most scientists think that Merkel, a chemist by training, will continue the science-friendly course she has pursued since coming to power in 2005. Federal science budgets increased by 10% this year, to around €10.2 billion (US\$15 billion), the latest rise in a steady upward trend in science spending over the decade (see graphic).

"It is gratifying to see that politicians have a great amount of trust in science," says Matthias Kleiner, president of the DFG, Germany's main granting agency. "We do know that this means an obligation for us to work hard to deliver what society expects from science."

Germany spends slightly more than 2.5% of its gross domestic product on science and technology, putting it ninth among the world's most research-intensive nations. In June, the federal government and the prime ministers of Germany's 16 states agreed an extra €18-billion package for universities and research

"We are set to increase the general freedom to research, and ease restrictions to stem-cell research and genetic engineering."



organizations (see <http://tinyurl.com/l768c6>). The heart of the package is a continuance of the so-called 'excellence initiative', a competition launched by the previous Social Democrat-Green government to raise the profile of German universities (see 'University excellence initiative takes off'). Without that funding, the

J. MACDOUGALL/AFP/GETTY

Visions of a green future

Renewable energy now accounts for 17% of electricity generated in Germany — up from 6.3% a decade ago. "Germany is the only country in the world that has managed to substantially increase the share of renewable energy in a relatively short time," says Claudia Kemfert, head of climate, energy and transport at the German Institute of Economic Research in Berlin.

The reason is the Renewable Energy Sources Act (EEG), which most people credit with making Germany the world leader in renewable-energy generation. In the past four years, more than €9 billion (US\$13 billion) has been invested in new renewable-energy installations. The industry now

employs more than 200,000 people.

Introduced in 2000 by the previous Social Democrat-Green government, the law permits homeowners and farmers to connect solar power and biogas systems to the electricity grid. It also guarantees them a fixed price for the excess energy they produce and sell on the market; this year, these 'feed-in tariffs' for wind energy, hydroelectricity and geothermal energy were increased by another 10%. The law also requires businesses to buy energy generated from renewable sources first, in an amount proportional to the share of electricity generated nationally from renewables.

German companies lead the worldwide market for environmental technologies. But competitors, mainly in the United States, China and Japan, are catching up quickly. The administration of US President Barack Obama, for example, says it will invest hundreds of billions of dollars in solar, wind, biofuels and geothermal power, with a view to generating 25% of domestic energy from renewable sources by 2025. Last year, the United States surpassed Germany in installed wind-energy capacity, and China's wind-power capacity is also growing by leaps and bounds.

In solar energy, Japan and Germany are now almost on a par

in the percentage of electricity they generate from photovoltaic panels, and Arab states in the Middle East and North Africa are expected to significantly increase their solar-power capacities in the near future. "If we don't want to risk our global leadership, the government must maintain its support of the whole field," says Kemfert.

In particular, she argues, the new government will need to maintain the substance of the EEG, and create a market environment even more favourable for renewable energies. And she argues that the country's electricity grid must be expanded as quickly as possible — perhaps even under the aegis of a Europe-wide authority. **Q.S.**



VOLCANOES STIRRED BY CLIMATE CHANGE

Impact of global warming on natural disasters 'poorly understood'.

www.nature.com/news

GETTY



Angela Merkel has spearheaded large increases in federal science budgets.

initiative would have expired in 2011.

Biotechnology, environmental technology, energy and health research have also benefited from the German government's €14-billion high-tech strategy, which began in 2006. And thanks to massive public support, Germany has become a global market leader in 'green' technologies, such as wind turbines and photovoltaic cells (see 'Visions of a green future'). In August, the cabinet approved a plan to put 1 million electric vehicles on the road by 2020, including €500 million for developing novel battery technologies and building a network of charging stations.

Reducing red tape

If the Free Democrats, led by Guido Westerwelle, succeed in becoming the new coalition partner, they may use their influence to reduce red tape and restrictions in ethically sensitive branches of science.

"The economic situation won't allow excessive increases in science budgets," says Ulrike Flach, the party's spokeswoman for science and technology. "But we are set to increase the general freedom to research, and ease existing restrictions to stem-cell research and genetic engineering."

For instance, German researchers are

University excellence initiative takes off

"I've been warned that Germany is a country ruled by pessimism and bureaucracy," says Christopher Barner-Kowollik, a polymer researcher at the Karlsruhe Institute of Technology (KIT). "I found that none of this is true any more."

Last year, Barner-Kowollik moved his eight-strong group from the University of New South Wales in Sydney, Australia, to the KIT, one of nine winners in the first two rounds of the 'excellence initiative', launched in 2005, that has revitalized Germany's academic landscape. "No other recent programme had a similar impact on German science," says Matthias Kleiner, president of the DFG, Germany's leading grant agency.

The competition, which involves extra annual funding of around €30 million (US\$44 million), has over the past three years created some 4,200 new jobs in science. The largest group of beneficiaries are young group leaders, postdocs and PhD students. The initiative has also created hundreds of permanent positions for full professors and science managers.

About a quarter of the new posts have been filled by foreign scientists or by German researchers returning from abroad, and almost 40% are held by women.

The initiative has put an end to the decades-old doctrine that all German universities are equal. Applicants can apply for extra funding for

graduate schools, large-scale research cooperations called 'clusters of excellence' and for so-called 'future concepts'. Universities that succeed in all three lines of funding can call themselves 'elite'.

Institutions that have gained elite status so far include the KIT, the two universities in Munich, and universities in Aachen, Freiburg, Heidelberg, Göttingen, Konstanz and Berlin.

For the next round of funding, from 2012 to 2017, the money available will be increased to €2.7 billion. More than 100 universities have been invited to submit proposals for new research projects by autumn 2010. A panel of scientists and politicians will announce the winners in summer 2012. **Q.S.**

currently allowed to use only human embryonic stem-cell lines created before 1 May 2007. The DFG and most stem-cell researchers say these are sufficient, but Flach says that if the community feels more lines are needed, the Free Democrats will push to change or abolish that cut-off date.

In the past, the Free Democrats have also acted to release universities from ministerial restraint. In Bavaria and North Rhine-Westphalia, the two largest German states, Free Democrat science ministers have in recent years introduced sweeping reforms such as allowing universities to hire professors without ministry approval.

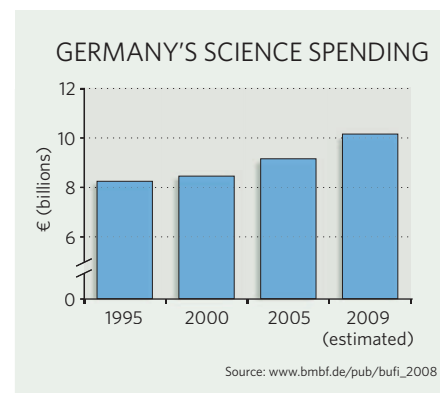
If Merkel's proposed coalition forms after the election, the post of science minister —

currently held by Annette Schavan of the Christian Democrats — might go to a Free Democrat. Flach says they will argue that the science ministry should take charge of technology-related budgets, including space and Germany's contribution to Galileo, the multi-billion-euro European global positioning system that is currently the responsibility of Germany's transport ministry.

But money remains the main concern of many researchers. Germany's biotechnology industry fears that a Christian Democrat-Free Democrat government might reduce support for high-tech industries in favour of more traditional technologies such as automation. "There is no wave of insolvency yet," says Viola Bronsema, managing director of Berlin-based BIO Deutschland, an association of Germany's small biotechnology companies. "But government support will be vital to keep the young industry alive."

The Greens also argue that support for environmental technologies and research into renewable energies must not fall behind. "There is still too much support for controversial branches, such as plant genetic engineering and nuclear research, which promise little benefit to society," says Hans-Josef Fell, the Green party's spokesman for energy and technology.

Germany's target for reducing annual greenhouse gases, by 40% of 1990 levels by 2020, is likely to remain unchanged.



Genomics shifts focus to rare diseases

COLD SPRING HARBOR, NEW YORK

Genome sequencing may finally be living up to its promise of pinpointing genetic mutations that bear on treatment for individual patients. But the breakthroughs are not coming from the DNA analysis of common diseases with complex genetic origins, which has been the obsession of genomics for nearly the past decade. Instead, many genome scientists are turning back to study rare disorders that are traceable to defects in single genes, and whose causes have remained a mystery.

The change is partly a result of frustration with the disappointing results of genome-wide association studies (GWAS). Rather than sequencing whole genomes, GWAS studies examine a subset of DNA variants in thousands of unrelated people with common diseases. Now, however, sequencing costs are dropping, and whole genome sequences can quickly provide in-depth information about individuals, enabling scientists to locate genetic mutations that underlie rare diseases by sequencing a handful of people.

"Years ago, people were using families and mapping approaches to distil down to a region where they thought a causative gene was," says Elaine Mardis, a director of the Genome Sequencing Center at Washington University in St Louis, Missouri. "Fast-forward 12 years, and you've got sort of the same thing going on, except with new technology that gives us much higher resolution and speed."

The change was showcased at the second Personal Genomes meeting, held in Cold Spring Harbor, New York, last week. At the same meeting last year, most speakers focused on the genome of scientist James Watson — one of only four fully sequenced individuals available at that time (see *Nature* 455, 1014; 2008). Now, about 50 individual genomes are published or in production, many of mostly anonymous patients with medical needs, estimates Richard Gibbs, director of the Human Genome Sequencing Center at Baylor College of Medicine in Houston, Texas.

The first genomes of patients came in cancer, but scientists are now quickly moving into less high-profile diseases. For instance, together with Baylor's James Lupski, Gibbs reported that they had sequenced the genomes of patients in one family with a familial neuropathy — a disorder marked by muscle weakness and pain — and found defects in a single gene that could account

"I now think that we're going to get there by understanding a whole lot of these Mendelian diseases."



Hugh Rienhoff sequenced family transcriptomes to try to diagnose his daughter Bea's genetic disease.

for the family's condition. Matthew Bainbridge, a student in Gibbs's lab, reported that he had found a genetic glitch that is probably responsible for an inherited form of ataxia, a disorder affecting bodily coordination, by sequencing the exome, or set of all protein-coding genes, of two distant relatives with the disease. And Jay Shendure of the University of Washington in Seattle used a separate exome-sequencing strategy to find the gene that could be responsible for Miller syndrome, marked by head and facial abnormalities.

Meanwhile, Richard Lifton of Yale University in New Haven, Connecticut, reported a striking example of how genome sequencing can help patients. A doctor asked Lifton to study a sickly infant who appeared to have a kidney disease. But Lifton's group sequenced the exomes of the infant and some family members and found a genetic variant in a gene, called SLC26A3, that causes congenital chloride diarrhoea, a treatable disease. Lifton informed the doctor, who reported that, indeed, the infant had had hourly bouts of diarrhoea.

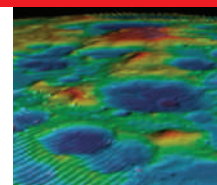
The meeting also heard from bioentrepre-

neur Hugh Rienhoff, a California father who has sequenced his daughter's transcriptome — the readout of protein templates expressed in her cells — as well as his wife's and his own to find the cause of her undiagnosed genetic disease (see <http://tinyurl.com/mf6oxs>). The effort has yielded a list of mutated genes that could cause her unique collection of symptoms, including long hands and feet and a cleft uvula, says Rienhoff, who trained as a clinical geneticist.

He argues that studies like his are promising first steps to understanding more complex diseases. "First, let's figure out the diseases that are 100% genetic and then go after the diseases that aren't 100% genetic," he says, pointing out that there are about 3,000 Mendelian diseases — inherited disorders caused by defects in single genes — of which the genetic causes are not known. "It's the new new thing, which is old: studying the rare stuff because it bears on the common stuff," he says.

"The issue is how we're going to understand the architecture of common disorders," agrees Gibbs. "Three or four years ago I thought it was going to be because of GWAS studies, but I now think that we're going to get there by understanding a whole lot of these Mendelian diseases."

H. RIENHOFF



WATER ON THE MOON?
Lunar missions find evidence of ice and hydrated minerals.
www.nature.com/news

NASA

Leslie Biesecker, who heads the ClinSeq study at the National Human Genome Research Institute in Bethesda, Maryland, is uncovering evidence to support that idea. ClinSeq has enrolled 725 patients and sequenced 251 of their genes, including two genes that, when mutated, cause familial hypercholesterolaemia (FH), a disease marked by high cholesterol levels. In seven patients with high cholesterol, Biesecker found mutations in the two FH genes, even though some of the patients had not been diagnosed with the condition. Biesecker estimates that 75–100 family members of these patients probably have undiagnosed FH, and is trying to contact them.

“If we can leverage the finding to diagnose and treat 75–100 people for FH, I can go to bed at night feeling like our sequencing money has been well spent,” he says.

Plummeting prices

Cost is still a roadblock to broader sequencing, however, so scientists at the meeting were interested in a report from collaborators of Complete Genomics, of Mountain View, California, which says it will dramatically cut the price of sequencing by offering human-genome sequences for US\$5,000 next year. The company sequenced a family of four people for \$20,000 apiece, including two children with undiagnosed symptoms, for the Institute for Systems Biology in Seattle, Washington, whose founder, Lee Hood, is on Complete Genomics’ scientific advisory board. Lee Rowen, a researcher at the institute, said that its analysis of Complete Genomics’ data has yielded three candidate genes for the children’s disorders. “We were happy with what they gave us,” Rowen says.

However, she also estimates that there could be roughly 10–20 errors per megabase of DNA, which would translate to tens of thousands of errors per genome. The group is resequencing each error to try to pinpoint the error rate more closely.

That left many in the audience unsatisfied. “What use is a \$20,000 genome if you have to spend even more money to figure out if it’s right or not?” Mardis said.

Some contenders in the crowded sequencing field hope to ultimately push costs as low as hundreds of dollars per genome. Until then, the ultimate payoff on the promise of genomics — a true understanding of common diseases — remains over the horizon. But for now, a down payment appears to have been made. ■

Erika Check Hayden

The elephant and the neutrino

India’s environment minister Jairam Ramesh will visit the site of a proposed underground neutrino laboratory next month, to try to break the impasse between physicists and environmentalists over its construction.

The US\$160-million India-based Neutrino Observatory (INO) was to have been completed by 2012 to study the elusive particles known as neutrinos (see *Nature* 450, 13; 2007). But its construction is mired in controversy over the wisdom of locating the facility in prime elephant and tiger habitat at Singara in the Nilgiri Biosphere Reserve, 250 kilometres south of Bangalore.

The observatory applied for permission to begin construction at the Singara site in 2006; “there has been no reply to date,” says project spokesman Naba Mondal, a physicist at the Tata Institute of Fundamental Research in Mumbai. “All I know is we have not cleared it,” says A. S. Balanathan, principal chief conservator of forest for the state of Tamil Nadu, who declined to comment further.

Last month, 11 leading physicists, including Nobel laureates Sheldon Glashow and Masatoshi Koshihara, wrote to India’s prime minister Manmohan Singh urging that the project move forward. “The INO will bring more big science to India and enhance India’s role as an important player in front-line science,” they wrote. Meanwhile, prominent Indian conservationists are circulating and signing a letter laying out their concerns and asking that the observatory be sited elsewhere.

The Nilgiri reserve includes more than 5,500 square kilometres of continuous forest cover and six protected areas. The proposed location for the INO comes as close as 7 kilometres to the edge of one of the sanctuaries. The project involves digging out a 120-metre-long cavern at the end of a 2-kilometre-long tunnel inside a mountain. The cavern will house a magnetized iron calorimeter to detect the muons that are produced occasionally when neutrinos interact with matter.

The controversy stems from disagreements over the impact of the tunnelling and the increased human population on the fragile ecosystem. “Transporting the estimated 630,000 tonnes of debris and 147,000 tonnes of construction material would require about 156,000 truck trips through 35 kilometres of forest — and two tiger reserves,” says the NBR Alliance, a group of Indian organizations

concerned about the reserve. This means 468,000 hours of disturbance to animal movement routes, the alliance estimates.

The INO team “could hardly have picked a site in India more likely to damage wildlife,” says John Seed, an Australian environmentalist who has researched elephant habitats in India. “As well as being home to the largest single population of Asian elephants in the world,” he says, “the Nilgiri is also one of the most important tiger habitats in the country.”

Mondal disputes the tally of construction debris and says that the project will limit the number of daily truck trips and restrict them to daytime. But Priya Davidar, an ecologist at Pondicherry University, says that the environmental impact assessment the project submitted to state officials is seriously flawed. Davidar is president of the Association for Tropical Biology and Conservation, based in Washington DC, which passed its own resolution urging the Indian government



Big hurdles for India’s neutrino-detection lab.

not to permit construction and to look for an alternative site.

Davidar criticizes the project for limiting its search to only two sites; a better location, she says, would have been the Kolar gold mine in neighbouring Karnataka state, used for neutrino detection in 1965. But the Kolar mine is now closed and filled with water, and is not suitable for lowering heavy materials down, says Mondal. He says that after considering other sites, his team, along with the Geological Survey of India, identified Singara as “the best available site for locating the INO, based on safety, seismicity, as well as year-round accessibility.”

The minister will visit on 10 October. If a construction permit is denied, INO may have to start looking for another site. ■

Killugudi Jayaraman

T. & P. LEESON/ARDEA.COM



Scaling up production from the flask to commercial quantities of biofuels from algae is the real challenge.

D. J. PHILLIP/A/P/PA

Gold rush for algae

The second of four weekly articles on biofuels describes how oil giants and others are placing their bets on algae.

THE BUSINESS OF BIOFUELS

No longer lowly pond scum, algae have rocketed in status to what some say is the most promising 'green' fuel source of the future. With the likes of Bill Gates, the US military and ExxonMobil trumpeting their potential, "it's hard to find someone on the sidewalk in New York who hasn't heard about the idea of using algae for energy," says Harrison Dillon, president and chief technology officer of Solazyme, a biotechnology company in South San Francisco, California. The company this month signed an \$8.5-million deal to produce commercial quantities of algal fuel for the US Navy.

Algae's photosynthetic cells produce an oily goo, including various oils and ethanol, that can be converted into advanced biofuels. Since 2007, more than \$1 billion has been injected into algae-to-energy research and development, says Will Thurmond, president of Emerging Markets Online, an energy consulting firm in Houston, Texas. "The validation is good for us," says Tim Zenk, vice-president of corporate affairs at Sapphire Energy, an algal-biotech company based in San Diego, California, that has investment backing from philanthropist Gates.

Algae have several key traits that make them a desirable energy source. They can be grown on non-agricultural land in a fraction of the area required by conventional oil crops

such as maize (corn), soybean and palm. In addition, algae capture carbon dioxide and can thrive in domestic waste water or salt water. But experts warn that there are still high hurdles to overcome before algal biofuels can compete economically with conventional fossil fuels. Challenges include finding strains of algae that reliably produce high yields, keeping contamination at bay, developing cost-effective growth chambers and efficiently harvesting oil from the cells. "In the end, it's all going to come down to economics and what it's going to cost to produce this algal oil on a large, commercial scale on a dollar-per-gallon basis," says Al Darzins, who leads the algal biofuels programme at the National Renewable Energy Laboratory (NREL) in Golden, Colorado.

In 2007, the United States consumed some 150 billion litres of diesel. Algal-biofuel companies may be able to produce millions of litres of oil in the next few years, but probably not billions, says Darzins. "We don't have the infrastructure."

To date, the biggest investment boon for algae has come from oil giant ExxonMobil, which in July announced that it will invest \$600 million over five to six years in a partnership with Synthetic Genomics, a company in La Jolla,

California, co-founded by genomics pioneer J. Craig Venter. ExxonMobil has said that its investment is contingent on Synthetic Genomics achieving certain milestones, and that if their efforts are successful, it expects to spend billions more on final development and early commercialization of algal biofuels. Synthetic Genomics is in the early stages of its venture, but the company maintains that it has engineered algal

cells that can directly secrete hydrocarbons in pure form, in contrast to the standard process of splitting the cells open to harvest their oils.

Other major oil players that have entered the algal arena include Chevron, which partnered the NREL in 2007 and

helped to restart, with an undisclosed amount, a research programme that had been shuttered since 1996. Chevron also announced a deal with Solazyme in January 2008, although the partners will offer only vague details about the nature of their collaboration. And this August, BP announced a \$10-million algal investment in Martek Biosciences, a nutritional-science company based in Columbia, Maryland.

In addition, Dow Chemical Company is backing Algenol Biofuels, a company in Bonita Springs, Florida, that specializes in ethanol production from algae, in its quest for a \$25-million grant from the US Department of Energy.

"Algal-biofuel companies may be able to produce millions of litres of oil in the next few years, but probably not billions."

If Algenol gets the grant, it will construct a pilot plant with Dow at Dow's manufacturing site in Freeport, Texas, with the goal of capturing industrial carbon dioxide and producing alga-derived ethanol to generate ethylene, a building block for plastics.

Meanwhile, Sapphire Energy has garnered more than \$100 million from bigwig investors, including Gates's Cascade Investments and the Rockefeller family's venture-capital firm Venrock. Sapphire is using genetic engineering to boost several algal traits, including improved protection from predators and low-cost harvestability. It is also working to genetically manipulate the algae to produce oils that are nearly identical to crude oil as extracted from the ground.

And Solazyme's contract with the Navy is the first contract anywhere to manufacture commercial-scale quantities of next-generation biofuels. The contract requires that Solazyme deliver some 75,000 litres of F-76 renewable fuel, which is similar in composition to diesel fuel, over the next year. "This really raises the bar in what constitutes a true production capability versus an interesting research direction," says Dillon.

Still, many challenges remain. In May GreenFuel Technologies, a front-runner on the algal scene that had amassed some \$70 million in investments since 2001, announced that it was closing down. Sam Jaffe, an energy analyst with IDC Energy Insights, a research and analysis firm based in Framingham, Massachusetts, says that GreenFuel pursued too many different technologies, including expensive greenhouses to control algal growth conditions. "Growing algae is easy," says Jaffe. "Growing it as a business and making money off of it is about getting the costs down."

One of the biggest challenges is to reproduce laboratory conditions on a large scale. In the lab, it can be easier to control algal growth and to find strains that produce copious amounts of oil. "But it's a totally different story when you take this organism that behaves well in the laboratory and you put it in acres' worth of outdoor ponds," says Darzins. For this reason, some companies have opted to grow their algae in enclosed 'bioreactors'. But the costs of building bioreactors can be prohibitively expensive. The algae community is "still torn" between open ponds and closed bioreactors, Darzins says.

With so much enthusiasm and investor interest in algal technology, new companies have sprung up almost overnight. Some experts say that because much of the science behind these technologies is not peer reviewed and is done through privately held companies, it can be difficult to gauge their progress. "On the one

hand you get their hype, and on the other hand they're guarding everything so closely that you can't evaluate it," says Martha Groom, a conservation biologist at the University of Washington in Bothell. "I find that fairly frustrating."

Experts say that a few companies have made questionable assertions about how much fuel they can reap from their algae. "Unfortunately, a lot of people tout these technologies and yet don't have the production data to back it up," says Doug Henston, chief executive of Solix Biofuels, a renewable-energy company based in Fort Collins, Colorado, that opened an algal oil-production demonstration facility in July at a coal-bed methane plant in southwestern Colorado. "That's the unfortunate case because it clouds the picture and builds unrealistic expectations," he says. Solix hopes to push its production capacity from its current rate of about 14,000 litres per hectare per year to between 37,000 and 47,000 litres per hectare per year. However, some start-ups have claimed that they can reach oil-production capacities



Fuel source of the future?

as high as 900,000 litres per hectare per year, which, says Henston, is "thermodynamically impossible".

Dillon, of Solazyme, says that the recent involvement by big-league investors, oil giants and the US military will help sort out approaches that are leading somewhere from those that aren't. "I think it's a good thing that we've got some real expectations coming on," he says. "There's been a lot of hype. That has a time window on it, and that type of time window tends to close when major players with real expectations start getting involved." ■

Amanda Leigh Mascarelli

NEXT WEEK: CELLULOSIC ETHANOL

Correction

The Editorial 'Data's shameful neglect' (*Nature* 461, 145; 2009) stated that the Joint Information Systems Committee was established by the seven UK research councils. It was, in fact, established by the three Higher Education Funding Councils.

The edge of physics

Canada's Perimeter Institute of Theoretical Physics was intended to become a world leader in the field. **Eric Hand** finds out if it has lived up to its ambitions.

Working at the Perimeter Institute for Theoretical Physics comes with certain perquisites. Whenever recruits arrive at the Toronto airport, for example, they are met by a limousine and driven west along Canada's Route 401 into the rich farmlands of Ontario. Eighty-five kilometres later, the limousine works its way through the streets of the town of Waterloo, and lets them out in front of a sleek building of black, green and glass squares that stands next to a pond in Waterloo Park. Stepping inside, the recruits find wall-to-wall blackboards, working fireplaces, a sauna, multiple dispensers of free coffee and the Black Hole Bistro, which serves free lunches on Wednesdays. And if they say yes to the recruiting pitch, they get a free BlackBerry smart phone — plus the power, even as a postdoc, to invite collaborators for visits of up to 18 weeks in the year.

The building itself is a gleaming architectural marvel rising from the site of a former municipal ice-hockey rink, which had to be demolished to make way. The significance is not lost on Perimeter's director, Neil Turok, a South African-born physicist (see 'A theoretical firebomber') who understands ice hockey's place in the Canadian psyche. The puck has been passed, says Turok. Waterloo will soon be known the world over for theoretical physics. "Part of what Perimeter represents in Canada is this search for self-confidence — the idea that, yes, we can do something better than anyone else," he says.

Besides, says Turok, "Mike thinks hockey is a complete waste of time". Mike Lazaridis, a Turkish-born engineer, is both the co-founder of Research in Motion — the Waterloo-based company that makes the BlackBerry — and a romantic about the transformative power of basic research. A decade ago he gave Can\$100 million (US\$95 million) to start Perimeter, and in June 2008, gave Can\$50 million more.

Lazaridis is hardly the first philanthropist to give money to science. For example, entrepreneur Fred Kavli stumped up US\$7.5 million to help found each of 15 research institutes — in astrophysics, theoretical physics, neuroscience and nanotechnology — that have been established in his name. Such gifts are generally tied to universities with existing facilities, but Lazaridis conjured up something where there was nothing.

Independent institutes aren't new either. In 1930, for example, siblings Louis Bamberger and Caroline Bamberger Fuld contributed US\$5 million from the sale

of their department store business — about US\$65 million today — to launch the Institute for Advanced Study (IAS) in Princeton, New Jersey. And ever since, academics have been trying to create similar sanctuaries of pure research, free from the constrictions of teaching, grant writing and university management. But where institutes such as the IAS have a reputation for being "stuffy", as Turok puts it, Perimeter is intended to be "more energetic,

more free-thinking and less project focused".

Turok, who became Perimeter's second permanent director in October 2008, has been making energetic use of the riches at his command. In July, the institute began driving piles into the ground for a Can\$30 million expansion that will double the institute's square footage. And, in a time of recession, Turok is hiring. He already has the largest population of theoretical physics postdocs in the world, with 44. And he is planning to more than double his full-time faculty from 12 to 25. Eventually, the building expansion will allow the current research staff of 85 to triple to 250, including visitors. Stephen Hawking, a former colleague of Turok at the University of Cambridge, UK, is scheduled to preside over a decadal celebration in mid-October, which will provide an excuse to kick-start a campaign aimed at doubling the endowment from its present size of Can\$200 million.

Grand ambitions

In the institute's new five-year plan, which starts in 2010, Turok writes that the overarching goal is to be the world's leading centre for theoretical physics. But the pursuit of that goal actually started in 1999, when an intense, verbal and philosophical dreamer named Howard Burton was still trying to figure out what to do with a freshly minted physics PhD from the University of Waterloo. The 34-year-old Burton didn't seem headed for an academic career. And he was unexcited by the prospect of donning a tie and cranking out financial algorithms in New York City, as so many physicists were doing at the time. On a whim

"What the Perimeter Institute represents in Canada is the search for self-confidence."

— Neil Turok

S. FVSH



The gleaming facade of the Perimeter Institute.

— and perhaps with a whiff of desperation — Burton sent a cover letter to the chief executive of Research in Motion with the tagline: “Please help save me from a lucrative career on Wall Street.”

Lazaridis e-mailed back, telling Burton he was welcome to a regular job at Research in Motion. But might he be interested in this other idea he was pondering?

Intrigued, Burton agreed to lunch at an Italian restaurant in a suburban Toronto strip mall. “I’m talking about doing something big,” Lazaridis told him, according to Burton’s account in his new book, *First Principles* (see page 477). Lazaridis passed Burton a number written on a napkin — the beginning and end of the salary negotiation phase, apparently. “On a tactical level, I think he was feeling me out,” Burton says in an interview from Lyons, France, where he moved with his family after leaving Perimeter in 2007. But Burton was also feeling out Lazaridis, making sure the chief executive wanted some sort of scientific think tank — not another research lab for Research in Motion.

Burton took the job. On his first day of work, he came up with the name ‘Perimeter’, inspired by a walk around the curving edge of Lake Ontario. Next, he started visiting other physics institutes, partly to begin recruitment and partly to absorb the lessons of each. He decided that he wanted Perimeter to have a resident faculty like the IAS. But he also wanted to emulate the nimbleness of the Kavli Institute for Theoretical Physics (KITP) at the University of California, Santa Barbara. So he followed the KITP strategy of staying *au courant* with the hottest

A theoretical firebomber

“I haven’t properly moved,” says Neil Turok in his sparse office at the Perimeter Institute for Theoretical Physics in Waterloo, Ontario, the sharp residual of a South African accent poking through his plummy English tones. “I just literally came here with a suitcase and a bicycle.”

That is easy to believe. In June, nearly nine months after he moved from the University of Cambridge, UK, to become Perimeter’s new director, Turok’s bookshelf contains just two books and a bag of coffee beans. He has had the walls painted a neon green; Turok doesn’t share the previous director’s preference for black. The only real ‘baggage’, as he puts it, that he brought to Canada are the twin obsessions that fill the room’s blackboard.

One of those obsessions is theoretical cosmology: half the board is covered in formulae related to his iconoclastic account of the Universe’s origin. The other, represented by the organizational bubbles and management flowcharts filling the rest of the board, is his stewardship of a growing system of mathematical institutes in Africa. Back in June, he was about to meet with Canadian development

agencies in Ottawa and tell them his most fervent wish: that the next Einstein should come from the developing world. “It’s all about access,” Turok says, “because I believe that there are very many hugely talented people in the world who never realize their potential.”

He knows how eager to learn students from the developing world can be — because he was one.

Turok was born in Johannesburg, South Africa, in 1958 to anti-apartheid activist parents. One of his earliest memories was of a nighttime police raid. As police ransacked his house, his older brother stood to the side and hugged a banned Miriam Makeba jazz record under his nightshirt. In 1962, his father went to prison after a failed attempt to firebomb a post office. Later, says Turok, his mother was imprisoned for putting up a demonstration poster. After his parents got out of jail, the family fled — first to Kenya, then Tanzania and finally to Britain, where his father sold volumes of the *Encyclopaedia Britannica* door to door. When Turok entered the University of Cambridge, he had none of the means of his wealthy

peers. But he had seen lions.

Turok credits his parents for passing on their intensity. He also inherited an instinct for going against accepted ways of thinking. With his ideas for an ‘ekpyrotic’ Universe that arose out of a fiery collision rather than a bang, Turok has been a bit of a firebomber himself (see *Nature* **458**, 820–824; 2009). But even as he has tried to tear down theories based on the conventional wisdom, he has also been building institutions up.

In 2003, he opened the African Institute for Mathematical Sciences (AIMS) in Cape Town, South Africa. He and his parents — who returned home to become members of parliament in the 1990s — bought a derelict hotel and turned it into a school. AIMS now offers an immersive 9-month course for top students across the continent. A second AIMS in Nigeria is about to open with separate funding. Three more — in Senegal, Ghana and Ethiopia — are being contemplated. Turok says Can\$25 million could support these three and the Cape Town AIMS for five years.

The idea for AIMS arose decades earlier, when Turok was teaching mathematics to villagers in Lesotho before he began university. He recalls speaking to a miner while watching a football match on a dirt pitch. “He said he only liked one thing in school: Shakespeare. Then he recited some. This to me was amazing. It made me realize that we all have this within us, this deeper side.” At AIMS, Turok wants an Einstein, not a Shakespeare, but the point is still the same. “It’s all about brilliant young people. There are some geniuses among them.”

E.H.



Neil Turok hopes to find the next Einstein in a developing country.

E. HAND

E-HAND

research via visitors that arrive for short-term programmes and workshops (see *Nature* doi:10.1038/news.2009.699; 2009).

Burton also wanted Perimeter's researchers to tackle overlooked niches of science in the manner of the Santa Fe Institute in New Mexico, which was founded by a group of Los Alamos physicists in 1984 to explore then-obscure notions of complex adaptive systems. So he looked for researchers who were interested in neglected areas of physics such as the root problems of quantum mechanics, a research area he called quantum foundations; and non-string theory approaches to quantum gravity, which seeks a unification of gravity and quantum mechanics.

By 2001, Burton's first recruits were at work in Waterloo's historic post office building.

In late 2004, his growing crew moved into their new home by the park. And by 2007, he felt he had achieved most of his goals. Burton stepped down as director in June of that year. He had never planned to stay forever, he says. But he could also feel a distinct chill in the atmosphere: he says that Lazaridis, the chairman of Perimeter's board, had stopped trusting and liking him — presumably because Lazaridis had learned that Burton was writing a book about his experiences. A Perimeter spokesman says that Burton's departure was a private matter and unconnected with a book.

Unique set-up

Whatever the reason, Burton still bristles over what he perceives as Lazaridis' excessive control over the institute. He concedes that Lazaridis has given more to Perimeter than has any other source. But he points out the inventor's largesse is nearly matched by more than Can\$100 million in federal and provincial funding. Even the city of Waterloo chipped in, giving Perimeter its park site. Why don't any of these institutions — or any physicists for that matter — have a seat on the board, asks Burton. "Is this a situation where, effectively, it is a rich man's toy?" Lazaridis has had an unusually strong hand in the management of Perimeter, agrees IAS astrophysicist Scott Tremaine, a Canadian who has served on Perimeter's scientific advisory committee. "The usual tradition is that you leave your hands off," he says.

Turok acknowledges that, when he was first offered the job, he thought it a bit strange that the board consisted of lawyers, businessmen and engineers. "Why isn't this governed by a group of scientists?" he asked. But he



Free coffee helps physicists such as Nima Arkani-Hamed (left) work all night in comfortable and relaxed surroundings.

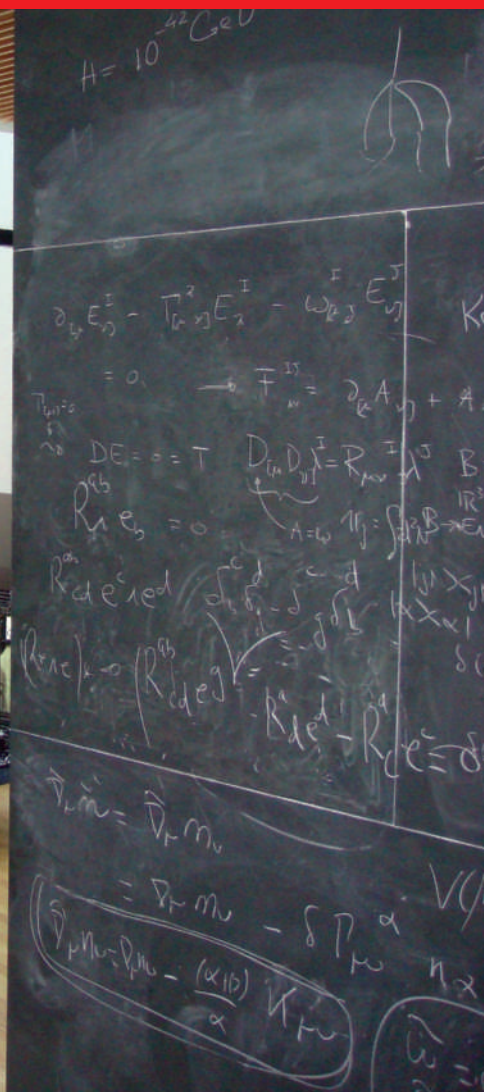
maintains that his relationship with Lazaridis, who declined to be interviewed by *Nature*, is fine — not least because Lazaridis has kept his promise never to interfere with Turok's authority on the strategy and scientific direction of the institute.

By now, says Turok, he actually prefers Perimeter's structure. The make up of the board helps give the place a risk-taking spirit that is more in keeping with a Silicon Valley start-up than an academic venture. And that means Turok can act faster. Since his arrival in October 2008, for example, he has doubled the areas on which researchers focus from four to eight, including a new emphasis on cosmology (see 'On the perimeter'). He has added a 9-month graduate programme to a place that had no teaching. And to remind staff that the coffers aren't bottomless, he has switched to a cheaper brand of the coffee that Perimeter provides for free.

This ability of Perimeter to move quickly — in matters both big and small — is what amazes Raymond LaFlamme, a quantum-information scientist who was plucked from Los Alamos National Laboratory in New Mexico as one of Perimeter's first hires. In 2002, he became director of a Perimeter spin-off: the Institute

for Quantum Computing (IQC) also in Waterloo, which has received a separate injection of nearly Can\$50 million from Lazaridis. The IQC has closer ties to the University of Waterloo, and therefore less independence than Perimeter, and LaFlamme has a telling example of the difference. He still keeps an office at Perimeter, and on a stand inside his office window is the receiver for a quantum-cryptography experiment that detects entangled photons from a transmitter on the University of Waterloo campus a few kilometres away. To allow the photons to reach the detector without destroying the quantum effects, LaFlamme needed to cut out a small piece of the tinted glass in his office window and replace it with clear glass. At a federally funded lab such as Los Alamos, he says, such a request would have been a nightmare. And at the nearby IQC building, where he keeps another receiver, the job took many months. At Perimeter, it took a day. "I went down the hall and talked to the guy in charge of the building. He said, 'Okay'. And — bang — it was done."

And yet — has this entrepreneurial energy led to new physics? Another bold goal of Turok's five-year plan is the expectation of "major scientific breakthroughs". "Everybody



On the perimeter

Neil Turok, director of the Perimeter Institute for Theoretical Physics in Waterloo, Ontario, has doubled the number of research areas that are focused on at the Canadian centre.

Quantum foundations

Concerns the still unanswered mysteries at the heart of quantum mechanics, the theory that describes the peculiar and probabilistic interactions of particles at very small scales.

Quantum information theory

Considers the processing of information held in quantum states that, unlike a classical bit, can exist simultaneously as both a 0 and a 1, and can be used to make unbreakable cryptographic codes.

Quantum gravity

Attempts to reconcile the forces of gravity called for by general relativity at cosmological scales, with quantum forces at very small scales.

Superstring theory

Describes a 'theory of everything' in which incredibly small strings create the fundamental forces of the Universe through their vibrations and interactions in 10 dimensions.

Particle physics

Concerns the interactions of particles at energy scales observed in astrophysical experiments, and created on Earth in accelerators such as the Large Hadron Collider at CERN, Europe's particle-physics laboratory near Geneva, Switzerland.

Cosmology

Considers questions such as: how was the Universe born? What were conditions like in the first seconds? What is the nature of dark matter? And why is the expansion of the Universe accelerating?

Condensed matter

Describes the behaviour of macroscopic materials under conditions achievable on Earth, such as superconductivity. But there are new areas of overlap with quantum mechanics and string theory.

Complex systems

A 'catch-all' category created by Turok to encompass research that lies at the intersection of other theoretical physics areas. This includes statistical physics, complex modelling and dynamic systems. **E.H.**



D. M. BENNETT/GETTY

Mike Lazaridis, head of the company behind BlackBerry phones, founded the Perimeter Institute.

wants a breakthrough," says KITP director and Nobel laureate David Gross, who is both a friend and mentor to Turok. But "you don't order results like that".

That's true, says Turok. But you can increase the odds by packing as much talent as possible into a room, and fuelling everyone with free coffee.

Different approach

The coffee is sometimes needed. Consider an evening in June, for example. The sun is sinking, but Nima Arkani-Hamed and Freddy Cachazo are just getting going. Several empty coffee mugs are scattered on Cachazo's desk.

Arkani-Hamed has just a day before he jets to Rome to present their work at a major string-theory conference, and they are feeling the pressure. Their hope is that, by dusting off a long forgotten corner of particle physics — S-Matrix theory, a model of particle interactions that began to be superseded by more fruitful accounts in the 1960s — they can clarify a very modern notion known as holography, which holds that information about the Universe can be encoded in fewer space-time dimensions than are apparent to us.

"Freddy and I are solving the mysteries of the Universe," Arkani-Hamed announces, only half-jokingly. The exuberant Arkani-Hamed sits at Cachazo's desk, while the more softly spoken Cachazo lounges on a sofa thumbing his standard-issue BlackBerry. One of Perimeter's young stars, Cachazo says he came to Canada from Venezuela partly to escape his country's heat. Arkani-Hamed, a 'distinguished research chair' at Perimeter, is a frequent visitor from the IAS. Keeping themselves going with a stash of cereal kept behind a stack of books, they work through the night, trying to determine the signs of 12 terms in an equation. They finally crack it

at 4 a.m., and leave an hour later for some sleep. It is a small step along what may end up being a blind alley. But it is the kind of effort that Turok wants: undirected, unconventional, ambitious.

Only time will tell whether these two physicists can reach their goal — or whether Perimeter itself will. "Early on," says Tremaine, "I was asked about Perimeter, and I said it might be the most important new institute since the IAS was founded," he says. "I still say that is true. And I still would use the word 'might'."

A few hours after Cachazo wakes up from his all-nighter, a delivery truck accidentally backs into the sprinkler system below Perimeter's car park. A fire alarm goes off. Dozens of physicists emerge, blinking in the hot, midsummer sun, to discover a dislodged sprinkler pipe spewing a jet of water. Half of its spray hits Perimeter's entrance wall, and the young physicists — some in shorts, some in khakis — happily

soak themselves in the mist. The other half of the jet soaks an Aston Martin belonging to Lazaridis, who is at Perimeter for one of his frequent visits. "Mike's getting a free car wash," says Turok, as Lazaridis ducks into his sports car.

The money man leaves, but the problems of physics remain. After the alarm is silenced, the physicists file back into the building's sun-soaked atrium, passing under a Greek inscription — the same phrase that supposedly hung from Plato's academy: "Let no one untrained in geometry enter here." Cachazo returns to his top-floor office, where the sun is beating in, and resumes work. But the air conditioning isn't working — perhaps because of the fire alarm. The temperature in his office climbs to 30°C Celsius. He feels sluggish. His brain is useless. The Venezuelan declares he is ready again for Waterloo's winter: "That's why I moved to Canada."

Eric Hand is a reporter for Nature based in Washington, DC.

"Is this effectively a rich man's toy?"
— Howard Burton

Cloudy, with a chance of science

When American and Chinese scientists agreed to measure pollution and dust over China, nobody foresaw how difficult it would be. **Jane Qiu** reports.

The meteorological bureau in the sleepy town of Shouxian in eastern China was buzzing with excitement. It was May 2008, and the spacious courtyard was littered with sophisticated remote-sensing instruments that had just arrived on loan from the United States Department of Energy (DoE). The bureau had been expecting the equipment earlier, but it had been held up by Chinese customs officials for more than two months.

A group of climate researchers and government officials from China and the United States eagerly inspected the new arrivals, which included a cloud radar, a tailor-made lidar (a radar-like instrument that sends out laser beams rather than microwaves) and sensors for studying various features in the atmosphere and the radiation from the Sun. “We can do great things with these here,” said Zhanqing Li, an atmospheric scientist at the University of Maryland at College Park, who was leading the Sino-American collaboration.

Over the next few months, these instruments would be pointed up into the Chinese sky to monitor and study aerosols — tiny airborne particles such as dust and soot. The researchers were particularly interested in tracing how aerosols alter the personality of clouds by influencing whether clouds produce rain, how high they extend, how much sunlight they reflect and

how long they persist. At present, atmospheric researchers have only a rudimentary understanding of how aerosols affect clouds and that ignorance is one of the major sources of uncertainty in forecasts of future climate.

For aerosol experts, China’s sky is close to heaven. The country has high concentrations of particles arising from pollution as well as natural dust blowing in from surrounding deserts. Researchers expected that data from such a particle-rich atmosphere would help to resolve major questions about aerosols and climate. At the same time, it was hoped that the project, staged at four sites across China (see map) would reap political rewards. The joint collaboration, conducted under the umbrella of the DoE’s Atmospheric Radiation

Measurements (ARM) programme, was viewed as a sign of China’s movement towards openness. “This kind of collaboration would have been inconceivable ten years ago,” says Thomas Ackerman, an atmospheric scientist at the University of Washington in Seattle.

The political winds did not, however, always blow favourably. With much frustration, the DoE had to alter its usual mode of operation and settle for lower-quality data and a smaller range of measurements than expected. “Between the heights of hope and the depths of despair, it was the most up-and-down deployment we have ever had,” says Warren Wiscombe, ARM’s chief scientist and an atmospheric scientist at NASA’s Goddard Space Flight Center in Greenbelt, Maryland.

The stakes are high because the data collected by the ARM programme will be used to improve the way that climate models simulate clouds and aerosols. When the programme was created in 1989, it collected measurements only at fixed sites within the United States. “But it soon dawned on us that we need as much information as possible from diverse climate systems around the world to build up a complete picture of global climate change,” says Wiscombe.

This resulted in a mobile facility, built in 2004, that contains most of the remote-sensing instruments present at the fixed sites. Each year,



D. GRAY/REUTERS

it visits a different region around the world, where it is run by an on-site technician. A team of scientists in the United States monitors each instrument remotely; they review the data for quality and then upload them in real time into the ARM data archive for use by the worldwide scientific community.

In 2007, Li and Chen Hongbin, deputy director of the Beijing-based Institute of Atmospheric Physics of the Chinese Academy of Sciences, won the bid to bring the mobile facility to China the following year. The project came under the umbrella of a climate-science agreement signed in 1987 between the DoE and China's science ministry. Under that compact, the two countries are committed to sharing data and collaborating on joint field campaigns, climate modelling and strategies for adapting to climate change.

Cloud puzzle

One aim of deploying the ARM mobile facility in China was to investigate an observation that had puzzled atmospheric scientists for some time. Microwave-sensing instruments on the joint US–Japanese Tropical Rainfall Measuring Mission satellite detect large amounts of liquid water in clouds over the coastal region of eastern China, yet the satellite's radar shows that there is very little rainfall.

"The two satellite instruments disagree with one another, which is very unusual," says Chris Kummerow, an atmospheric scientist at Colorado State University in Fort Collins, who discovered the discrepancy with his colleagues (W. Berg *et al.* *J. Appl. Meteorol.* **45**, 434–454; 2006). Some researchers suspect that the high level of aerosols in the Chinese atmosphere might be the culprit. The standard thinking about aerosols is that the particles often suppress rainfall by providing a nucleation site on which water can condense; they increase the number of cloud droplets and reduce their average size, thus making it harder for small droplets to grow big enough to drop out of the cloud as rain.

And China's air is chock-full of aerosols, at concentrations three times the global average. The density of aerosols, particularly in the populous eastern part of the country, has been rising rapidly. In eastern central China, Li and his colleagues have found that the number of clean air days — defined as having at least 75% of the maximum visibility — has declined continuously over the past three decades, from 26% in 1976 to 14% in 2007.

The full effect of all those aerosols is hard to discern. Many other variables, such as the types of aerosol, the nature of the land surface

and atmospheric circulation patterns, could influence how aerosols alter clouds. "We need a large data set from as many climate regions as possible," says Mark Miller, an atmospheric scientist at Rutgers University in New Brunswick, New Jersey. "The ARM deployment in China is an important part of that effort."

The plan was to take the monitoring facility to four locations in China with different climates and aerosol types. The two eastern sites were rural Shouxian, 500 kilometres northwest of Shanghai, and Taihu (Lake Tai), in the industrial heartland of the Yangtze River delta region, about 100 kilometres west of Shanghai. These



The arrival of the equipment in Shouxian was a cause for celebration.

spate of problems crushed their hopes. "At some points we thought we wouldn't be able to get anything out of it," he says.

The difficulties began right at the start, when the instruments became stuck in Chinese customs because of tightened regulations due to the Olympics. "We were not aware of some of the changes in customs regulations and did not have all

the paperwork necessary for getting the instruments through," says Chen. On top of that, some of the sophisticated, upward-pointing instruments, such as the cloud radar, raised much suspicion among customs officials and this helped hold up their release. The delay severely curtailed the field campaign, especially at Zhangye where many of the spring dust storms were missed.

When the instruments were finally up and running in China, the support team in the United States found that the Chinese government would not allow them to access the instruments through the Internet. They could not check the quality of the measurements or fix problems remotely, nor could they upload data to the public database. The on-site technician was able to solve most of the technical issues that emerged, but some subtle problems remained for much of the deployment in China.

To the DoE's further dismay, officials from the Chinese Meteorological Administration (CMA) shut down photometers at the Shouxian

At the end of the dust-storm season in July 2008, the instruments were moved to Xianghe to study how government actions to reduce pollution during the Beijing Olympics affected solar radiation reaching the surface. "You rarely have the opportunity to do something like that in a major metropolis," says Miller.

Stops and starts

Despite the high hopes, however, the work in China was an uphill struggle at each step, says operational manager Kim Nitschke. "Originally, our expectations were very high as we thought it would be an exciting time to get into China during the Olympics," says Wiscombe. Then, the



J. QIU; Z. LI; J. QIU

and Xianghe sites that were measuring particles of black carbon, a major component of soot. Interestingly, the CMA did not interfere with measurements at other observing sites, possibly because these were being jointly operated with Chinese universities.

At one point in July last year, the DoE threatened to terminate the ARM project in China, but it decided to stay on, in consideration of the effort already invested. Eventually, after lengthy negotiations, the American researchers were granted a two-week window in October 2008, during which they could connect with the instruments remotely from the United States for about four hours each day. Only then were they able to fix the broken cloud radar, enabling it to collect data during the last two months of the project.

Political problems caused headaches up to the final days of the year-long stay in China. "Even at the end of the deployment, we were not sure whether we would be able to get anything out of it," says Nitschke. "I had all the data with me on a portable hard drive, but wasn't sure whether I would manage to get them out of China or if the Chinese government would let us make them publicly available."

The CMA did give the green light, and the original data from Shouxian and Zhangye are now available from ARM's archive. Many researchers are philosophical about the difficulties encountered by ARM's China deployment. "China is going through a transitional phase," says Daniel Rosenfeld, an atmospheric scientist at the Hebrew University in Jerusalem, Israel. Although the country is still not as open as people would like it to be, the collaboration clearly signals a move away from the old ways, he says. "The process is not complete yet, but I hope the issue of openness will be a thing of the past soon." Wei-Chyung Wang, a climate scientist at the State University of New York at Albany, and the United States' chief scientist managing the climate-science agreement, says that the project is "one of the most successful collaborations under the agreement". He adds that "this unique experience has really opened up the dialogue and will stimulate more interest in similar collaborations".

Data feast

Participants in the project say they are happy with what they were able to collect. "We now have cloud data from China nobody has ever had before," says Miller. Since the data collected at Shouxian and Zhangye were made publicly available in March and April 2009,



Clockwise, a balloon used to measure dust; collaboration participants; and a collection of radiometers.



Q. JI, Z. LI: ARM CLIMATE RES. FACILITY

respectively, researchers have been busy analysing them. Connor Flynn, an atmospheric scientist at the DoE's Pacific Northwest National Laboratory in Richland, Washington, is excited by the data. "Some of the lidar images are just tragically breathtaking," he says of the pollution measurements.

He has plotted the lidar data to reveal the concentration and composition of aerosols at various altitudes. "You can see atmospheric layers swirling together, ice crystals falling from a high cloud and their properties changing as they go through other layers," says Flynn. The data also show air masses with distinct aerosol compositions coming together from dust, urban, industrial and agricultural regions and mixing at different altitudes. "You don't always have those diverse sources of aerosols at high concentrations in other parts of the world,"

he says. The information will be enormously valuable for climate modellers trying to simulate those processes, Flynn adds.

Using data from radiometers, lidar, cloud radar and weather balloon, Li and his colleagues have made some headway in understanding the water-rainfall paradox of the clouds over eastern China. At a meeting jointly convened in Beijing last month by the DoE and China's science ministry, Li showed that the effect of aerosols on rainfall depends on the amount of liquid water in the clouds. When clouds are relatively dry, adding aerosols can suppress precipitation. In wetter clouds, however, aerosols make rain more likely. This is consistent with the observation that the number of days of light rainfall has decreased by 23% in the past 50 years in eastern China; cloud modelling studies show that this can be explained by the

increased aerosol concentrations in the region (Y. Qian *et al.* *J. Geophys. Res.* 114, D00K02; 2009). What's more, data collected at Shouxian and Taihu show that aerosols apparently affect the thickness of clouds and the altitudes at which they form. "If this proves to be the case, the implications of aerosols for climate change will be tremendous," he says.

But the Chinese data have not solved all the cloud conundrums. Using a computer simulation, Kummerow and his colleagues have pinpointed "a very complicated pattern" in how aerosols affect clouds. "They seem to increase precipitation in some places and decrease rainfall in others," he says. For example, the cloud radar and radiometer data from Shouxian indicate that, on days with comparable aerosol amounts and cloud liquid water, sometimes it rained and sometimes it didn't.

These preliminary passes through the ARM data collected in China show how much remains to be learned there. And the scientific results are not the only dividends the project has yielded. To Chinese researchers such as Huang Jianping, an atmospheric scientist at Lanzhou University in Gansu province, the ARM collaboration helped strengthen China's capability to run long-term field campaigns. Since 2005, Huang has been building up an observational site 40 kilometres east of Zhangye, focusing on climate research of semi-arid regions. "The ARM deployment in Zhangye has allowed us to work alongside the best people in the field," says Huang. And that experience has helped spur Chinese scientists to set their sights far higher.

Jane Qiu writes for *Nature* from Beijing.

"We now have cloud data from China nobody has ever had before."
— Mark Miller

CORRESPONDENCE

Luxury bushmeat trade threatens lemur conservation

SIR — Shocking new proof of an emerging trade in lemur bushmeat in Madagascar (see <http://tinyurl.com/mqsx7w>) is refocusing attention on the conservation and health challenges in one of the world's most important biodiversity hotspots.

The growth of this market, in which lemurs are sold as a delicacy to luxury consumers, could mean extinction for already-endangered lemur species, which are found only in Madagascar. Furthermore, as in other countries, bushmeat hunting carries serious risks to public health by fostering emergence of disease.

Madagascar has experienced an upsurge in environmental crime since its political upheaval in March this year. Increasing illegal harvesting of precious hardwoods and animal trafficking bodes poorly for the future of Madagascar's already-degraded environment, where 90% of its original forest cover has been lost. Political chaos and the withdrawal of foreign aid mean that these environmental crimes have continued almost unchecked.

Russell Mittermeier, president of Conservation International, has warned that certain lemur species, such as the golden-crowned sifaka (*Propithecus tattersalli*), could vanish as a result of hunting for the new market. There may be as many as 99 lemur species in Madagascar (R. A. Mittermeier *et al.* *Intl. J. Primatol.* **29**, 1607–1656; 2008). Wiping out any of these would disrupt the ecological balance and undermine the country's ecotourism industry.

Some 75% of emerging diseases have zoonotic origins (L. H. Taylor *et al.* *Phil. Trans. R. Soc. Lond. B* **356**, 983–989; 2001). Ebola and simian foamy virus outbreaks, for example, as well as HIV, have been traced to bushmeat hunting and butchering. The increase in human–wildlife contact in

Madagascar's degraded forests, along with its extreme biodiversity and wide distribution of domestic animals, could enhance the risk of disease emergence and spread, potentially to a global level.

The country's interim government has responded to the crisis by firing several forestry officials, but more cohesive enforcement is needed. Mittermeier has urged the international community to reinstate conservation funding to Madagascar, in order to save this pinnacle of biodiversity. With 20% of the world's primate species in peril, and with increased risks of disease emergence, an integrated solution must and can be achieved by conservation, public health and development interests.

Meredith A. Barrett Box 90338,
University Program in Ecology,
Duke University, Durham,
North Carolina 27705, USA
e-mail: meredith.barrett@duke.edu
Jonah Ratsimbazafy Durrell Wildlife
Conservation Trust, BP 8511,
Antananarivo 101, Madagascar
e-mail: jonah.ratsimbazafy@durrell.org

A communication wipeout by gabbling presenters

SIR — I have noticed a trend among speakers at scientific conferences to speed up their oral presentations so that they can compress as much information as possible into their allocated time slots. Talking so fast can create a problem for those in the audience whose native language is not the one being used by the speaker — almost invariably English on today's stage.

For oral presentations to function as a clear mode of communication, I have some advice for speakers addressing an international audience. Bear in mind the diversity of your listeners and practise your talk on scientists whose first language is not English (and on those outside

your discipline, in the case of a broader-themed conference); keep to the speed of a public-radio news broadcast; avoid cluttering your speech with unnecessary words and jargon; and do not launch into an unrelated story while presenting a text slide.

Contrary to popular belief, text slides and slide reading are not always boring, particularly for people listening to a talk in an unfamiliar language. And spare a thought for foreigners in the audience who may find jokes confusing: make sure that the humour translates!

Dongwook Ko School of Environmental
and Life Sciences, Charles Darwin
University, Darwin, Northern Territory
0909, Australia
e-mail: d.ko@uq.edu.au

Politics and priorities behind Greek research reforms

SIR — In your News story 'Greek scientists fight research shake-up' (*Nature* **460**, 671; 2009), you indicate that research at the National Hellenic Research Foundation (NHRF) was judged to be generally poor during government-backed evaluations in 2001 and 2005. This is not the case.

On each occasion, three of the six NHRF institutes were top-rated and named as centres of excellence. At none of the other three did the evaluators find the research weak.

Among those that were top-rated in 2005, the NHRF's Institute of Biological Research and Biotechnology (IBRB), of which I am director, was graded highest of the biological research institutes in the Attica region (including the Fleming institutes) and the second-highest in the country. Yet the IBRB is earmarked by the government for dissolution and absorption by the Alexander Fleming Biomedical Sciences Research Centre.

Concerning plans to restructure research in Greece, the scientific

community is agreed that some changes are definitely required — including more financial investment, greater respect for the process of scientific evaluation and the establishment of national priorities. However, the latest plans were not presented for discussion, so they are widely viewed as a policy that was chosen mainly to satisfy electoral prospects and client interests — with catastrophic consequences for the future of research in Greece.

The IBRB-NHRF is contributing to the national dialogue on the matter by presenting a detailed proposal for changes that could benefit the development of Greek biological research (see www.eie.gr).

Fragiskos N. Kolis School of Chemical
Engineering, National Technical
University of Athens, and IBRB-NHRF,
48 Vassileos Constantinou Avenue,
116 35 Athens, Greece
e-mail: kolis@eie.gr

Overzealous use of decimal places has wrong kind of impact

SIR — Scientists teach students to evaluate critically the significance of their measurements, and to eschew meaningless decimal places thrown up when pocket calculators work out a quotient of two integers. So what are we to make of the recently released impact factors, including *Nature's* much advertised rating of '31.434' (see also www.nature.com/nature/about)? Has Thomson Reuters discovered a protocol that allows it to measure the impact of a journal with an accuracy of 32 p.p.m.?

Quoting this figure conveys the wrong impression — that innumerate marketing is trumping common sense at the heart of science's leading journal.

J. M. D. Coey School of Physics and
Centre for Research on Adaptive
Nanostructures and Nanodevices,
Trinity College, Dublin 2, Ireland
e-mail: jcoey@tcd.ie

FEATURE

A safe operating space for humanity

Identifying and quantifying planetary boundaries that must not be transgressed could help prevent human activities from causing unacceptable environmental change, argue **Johan Rockström** and colleagues.

Although Earth has undergone many periods of significant environmental change, the planet's environment has been unusually stable for the past 10,000 years^{1–3}. This period of stability — known to geologists as the Holocene — has seen human civilizations arise, develop and thrive. Such stability may now be under threat. Since the Industrial Revolution, a new era has arisen, the Anthropocene⁴, in which human actions have become the main driver of global environmental change⁵. This could see human activities push the Earth system outside the stable environmental state of the Holocene, with consequences that are detrimental or even catastrophic for large parts of the world.

During the Holocene, environmental change occurred naturally and Earth's regulatory capacity maintained the conditions that enabled human development. Regular temperatures, freshwater availability and biogeochemical flows all stayed within a relatively narrow range. Now, largely because of a rapidly growing reliance on fossil fuels and



SUMMARY

- New approach proposed for defining preconditions for human development
- Crossing certain biophysical thresholds could have disastrous consequences for humanity
- Three of nine interlinked planetary boundaries have already been overstepped

industrialized forms of agriculture, human activities have reached a level that could damage the systems that keep Earth in the desirable Holocene state. The result could be irreversible and, in some cases, abrupt environmental change, leading to a state less conducive to human development⁶. Without pressure from humans, the Holocene is expected to continue for at least several thousands of years⁷.

Planetary boundaries

To meet the challenge of maintaining the Holocene state, we propose a framework based on 'planetary boundaries'. These

boundaries define the safe operating space for humanity with respect to the Earth system and are associated with the planet's biophysical subsystems or processes. Although Earth's complex systems sometimes respond smoothly to changing pressures, it seems that this will prove to be the exception rather than the rule. Many subsystems of Earth react in a nonlinear, often abrupt, way, and are particularly sensitive around threshold levels of certain key variables. If these thresholds are crossed, then important subsystems, such as a monsoon system, could shift into a new state, often with deleterious or potentially even disastrous consequences for humans^{8,9}.

Most of these thresholds can be defined by a critical value for one or more control variables, such as carbon dioxide concentration. Not all processes or subsystems on Earth have well-defined thresholds, although human actions that undermine the resilience of such processes or subsystems — for example, land and water degradation — can increase the risk that thresholds will also be crossed in other processes, such as the climate system.

We have tried to identify the Earth-system processes and associated thresholds which, if crossed, could generate unacceptable environmental change. We have found nine such processes for which we believe it is necessary to define planetary boundaries: climate change; rate of biodiversity loss (terrestrial and marine); interference with the nitrogen and phosphorus cycles; stratospheric ozone depletion; ocean acidification; global freshwater use; change in land use; chemical pollution; and atmospheric aerosol loading (see Fig. 1 and Table).

In general, planetary boundaries are values for control variables that are either at a 'safe' distance from thresholds — for processes with evidence of threshold behaviour — or at dangerous levels — for processes without

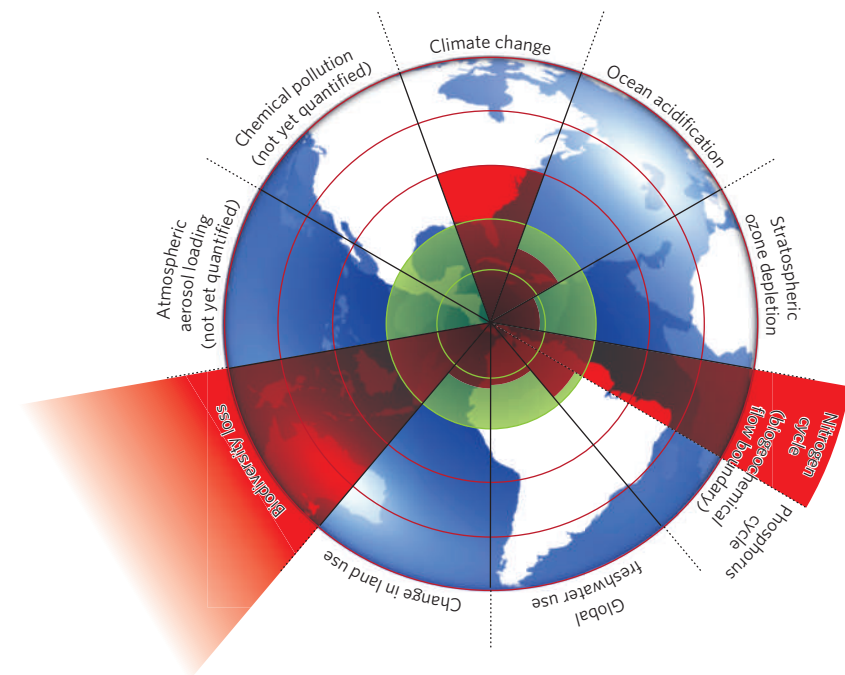


Figure 1 | Beyond the boundary. The inner green shading represents the proposed safe operating space for nine planetary systems. The red wedges represent an estimate of the current position for each variable. The boundaries in three systems (rate of biodiversity loss, climate change and human interference with the nitrogen cycle), have already been exceeded.

evidence of thresholds. Determining a safe distance involves normative judgements of how societies choose to deal with risk and uncertainty. We have taken a conservative, risk-averse approach to quantifying our planetary boundaries, taking into account the large uncertainties that surround the true position of many thresholds. (A detailed description of the boundaries — and the analyses behind them — is given in ref. 10.)

Humanity may soon be approaching the boundaries for global freshwater use, change in land use, ocean acidification and interference with the global phosphorous cycle (see Fig. 1). Our analysis suggests that three of the Earth-system processes — climate change, rate of biodiversity loss and interference with the nitrogen cycle — have already transgressed their boundaries. For the latter two of these, the control variables are the rate of species loss and the rate at which N_2 is removed from the atmosphere and converted to reactive nitrogen for human use, respectively. These are rates of change that cannot continue without significantly eroding the resilience of major components of Earth-system functioning. Here we describe these three processes.

Climate change

Anthropogenic climate change is now beyond dispute, and in the run-up to the climate negotiations in Copenhagen this December, the international discussions on targets for climate mitigation have intensified. There is a growing convergence towards a '2°C guard-rail' approach, that is, containing the rise in global mean temperature to no more than 2°C above the pre-industrial level.

Our proposed climate boundary is based on two critical thresholds that separate qualitatively different climate-system states. It has two parameters: atmospheric concentration of carbon dioxide and radiative forcing (the rate of energy change per unit area of the globe as measured at the top of the atmosphere). We propose that human changes to atmospheric CO_2 concentrations should not exceed 350 parts per million by volume, and that radiative forcing should not exceed 1 watt per square metre above pre-industrial levels. Transgressing these boundaries will increase the risk of irreversible climate change, such as the loss of major ice sheets, accelerated sea-level rise and abrupt shifts in forest and agricultural systems. Current CO_2 concentration stands at 387 p.p.m.v. and the change in radiative forcing is 1.5 W m^{-2} (ref. 11).

There are at least three reasons for our proposed climate boundary. First, current climate models may significantly underestimate the severity of long-term climate change for

PLANETARY BOUNDARIES

Earth-system process	Parameters	Proposed boundary	Current status	Pre-industrial value
Climate change	(i) Atmospheric carbon dioxide concentration (parts per million by volume)	350	387	280
	(ii) Change in radiative forcing (watts per metre squared)	1	1.5	0
Rate of biodiversity loss	Extinction rate (number of species per million species per year)	10	>100	0.1–1
Nitrogen cycle (part of a boundary with the phosphorus cycle)	Amount of N_2 removed from the atmosphere for human use (millions of tonnes per year)	35	121	0
Phosphorus cycle (part of a boundary with the nitrogen cycle)	Quantity of P flowing into the oceans (millions of tonnes per year)	11	8.5–9.5	–1
Stratospheric ozone depletion	Concentration of ozone (Dobson unit)	276	283	290
Ocean acidification	Global mean saturation state of aragonite in surface sea water	2.75	2.90	3.44
Global freshwater use	Consumption of freshwater by humans (km^3 per year)	4,000	2,600	415
Change in land use	Percentage of global land cover converted to cropland	15	11.7	Low
Atmospheric aerosol loading	Overall particulate concentration in the atmosphere, on a regional basis	To be determined		
Chemical pollution	For example, amount emitted to, or concentration of persistent organic pollutants, plastics, endocrine disruptors, heavy metals and nuclear waste in, the global environment, or the effects on ecosystem and functioning of Earth system thereof	To be determined		

Boundaries for processes in red have been crossed. Data sources: ref. 10 and supplementary information

a given concentration of greenhouse gases¹². Most models¹¹ suggest that a doubling in atmospheric CO_2 concentration will lead to a global temperature rise of about 3°C (with a probable uncertainty range of 2–4.5°C) once the climate has regained equilibrium. But these models do not include long-term reinforcing feedback processes that further warm the climate, such as decreases in the surface area of ice cover or changes in the distribution of vegetation. If these slow feedbacks are included, doubling CO_2 levels gives an eventual temperature increase of 6°C (with a probable uncertainty range of 4–8°C). This would threaten the ecological life-support systems that have developed in the late Quaternary environment, and would severely challenge the viability of contemporary human societies.

The second consideration is the stability of the large polar ice sheets. Palaeoclimate data from the past 100 million years show that CO_2 concentrations were a major factor in the long-term cooling of the past 50 million years. Moreover, the planet was largely ice-free until CO_2 concentrations fell below 450 p.p.m.v.

(± 100 p.p.m.v.), suggesting that there is a critical threshold between 350 and 550 p.p.m.v. (ref. 12). Our boundary of 350 p.p.m.v. aims to ensure the continued existence of the large polar ice sheets.

Third, we are beginning to see evidence that some of Earth's subsystems are already moving outside their stable Holocene state. This includes the rapid retreat of the summer sea ice in the Arctic ocean¹³, the retreat of mountain glaciers around the world¹¹, the loss of mass from the Greenland and West Antarctic ice sheets¹⁴ and the accelerating rates of sea-level rise during the past 10–15 years¹⁵.

Rate of biodiversity loss

Species extinction is a natural process, and would occur without human actions. However, biodiversity loss in the Anthropocene has accelerated massively. Species are becoming extinct at a rate that has not been seen since the last global mass-extinction event¹⁶.

The fossil record shows that the background extinction rate for marine life is 0.1–1 extinctions per million species per year; for

mammals it is 0.2–0.5 extinctions per million species per year¹⁶. Today, the rate of extinction of species is estimated to be 100 to 1,000 times more than what could be considered natural. As with climate change, human activities are the main cause of the acceleration. Changes in land use exert the most significant effect. These changes include the conversion of natural ecosystems into agriculture or into urban areas; changes in frequency, duration or magnitude of wildfires and similar disturbances; and the introduction of new species into land and freshwater environments¹⁷. The speed of climate change will become a more important driver of change in biodiversity this century, leading to an accelerating rate of species loss¹⁸. Up to 30% of all mammal, bird and amphibian species will be threatened with extinction this century¹⁹.

Biodiversity loss occurs at the local to regional level, but it can have pervasive effects on how the Earth system functions, and it interacts with several other planetary boundaries. For example, loss of biodiversity can increase the vulnerability of terrestrial and aquatic ecosystems to changes in climate and ocean acidity, thus reducing the safe boundary levels of these processes. There is growing understanding of the importance of functional biodiversity in preventing ecosystems from tipping into undesired states when they are disturbed²⁰. This means that apparent redundancy is required to maintain an ecosystem's resilience. Ecosystems that depend on a few or single species for critical functions are vulnerable to disturbances, such as disease, and at a greater risk of tipping into undesired states^{8,21}.

From an Earth-system perspective, setting a boundary for biodiversity is difficult. Although it is now accepted that a rich mix of species underpins the resilience of ecosystems^{20,21}, little is known quantitatively about how much and what kinds of biodiversity can be lost before this resilience is eroded²². This is particularly true at the scale of Earth as a whole, or for major subsystems such as the Borneo rainforests or the Amazon Basin. Ideally, a planetary boundary should capture the role of biodiversity in regulating the resilience of systems on Earth. Because science cannot yet provide such information at an aggregate level, we propose extinction rate as an alternative (but weaker) indicator. As a result, our suggested planetary boundary for biodiversity of ten times the background rates of extinction is only a very preliminary estimate. More research is required to pin down this boundary with greater certainty. However, we can say with some confidence that Earth cannot sustain the current rate of loss without significant erosion of ecosystem resilience.

Nitrogen and phosphorus cycles

Modern agriculture is a major cause of environmental pollution, including large-scale nitrogen- and phosphorus-induced environmental change²³. At the planetary scale, the additional amounts of nitrogen and phosphorus activated by humans are now so large that they significantly perturb the global cycles of these two important elements^{24,25}.

Human processes — primarily the manufacture of fertilizer for food production and the cultivation of leguminous crops — convert around 120 million tonnes of N₂ from the atmosphere per year into reactive forms — which is more than the combined effects from all Earth's terrestrial processes. Much of this new reactive nitrogen ends up in the environment, polluting waterways and the coastal zone, accumulating in land systems and adding a number of gases to the atmosphere. It slowly erodes the resilience of important Earth subsystems. Nitrous oxide, for example, is one of the most important non-CO₂ greenhouse gases and thus directly increases radiative forcing.

Anthropogenic distortion of the nitrogen cycle and phosphorus flows has shifted the state of lake systems from clear to turbid water²⁶. Marine ecosystems have been subject to similar shifts, for example, during periods of anoxia in the Baltic Sea caused by excessive nutrients²⁷. These and other nutrient-generated impacts justify the formulation of a planetary boundary for nitrogen and phosphorus flows, which we propose should be kept together as one boundary given their close interactions with other Earth-system processes.

Setting a planetary boundary for human modification of the nitrogen cycle is not straightforward. We have defined the boundary by considering the human fixation of N₂ from the atmosphere as a giant 'valve' that controls a massive flow of new reactive nitrogen into Earth. As a first guess, we suggest that this valve should contain the flow of new reactive nitrogen to 25% of its current value, or about 35 million tonnes of nitrogen per year. Given the implications of trying to reach this target, much more research and synthesis of information is required to determine a more informed boundary.

Unlike nitrogen, phosphorus is a fossil mineral that accumulates as a result of geological processes. It is mined from rock and its uses range from fertilizers to toothpaste. Some 20 million tonnes of phosphorus is mined every year and around 8.5 million–9.5 million tonnes of it finds its way into the oceans^{25,28}. This is estimated to be approximately eight times the natural background rate of influx.

Records of Earth history show that large-scale ocean anoxic events occur when critical thresholds of phosphorus inflow to the oceans are crossed. This potentially explains past mass extinctions of marine life. Modelling suggests that a sustained increase of phosphorus flowing into the oceans exceeding 20% of the natural background weathering was enough to induce past ocean anoxic events²⁹.

Our tentative modelling estimates suggest that if there is a greater than tenfold increase in phosphorus flowing into the oceans (compared with pre-industrial levels), then anoxic ocean events become more likely within 1,000 years. Despite the large uncertainties involved, the state of current science and the present observations of abrupt phosphorus-induced regional anoxic events indicate that no more than 11 million tonnes of phosphorus per year should be allowed to flow into the oceans — ten times the natural background rate. We estimate that this boundary level will allow humanity to safely steer away from the risk of ocean anoxic events for more than 1,000 years, acknowledging that current levels already exceed critical thresholds for many estuaries and freshwater systems.

Delicate balance

Although the planetary boundaries are described in terms of individual quantities and separate processes, the boundaries are tightly coupled. We do not have the luxury of concentrating our efforts on any one of them in isolation from the others. If one boundary is transgressed, then other boundaries are also under serious risk. For instance, significant land-use changes in the Amazon could influence water resources as far away as Tibet³⁰. The climate-change boundary depends on staying on the safe side of the freshwater, land, aerosol, nitrogen–phosphorus, ocean and stratospheric boundaries. Transgressing the nitrogen–phosphorus boundary can erode the resilience of some marine ecosystems, potentially reducing their capacity to absorb CO₂ and thus affecting the climate boundary.

The boundaries we propose represent a new approach to defining biophysical preconditions for human development. For the first time, we are trying to quantify the safe limits outside of which the Earth system cannot continue to function in a stable, Holocene-like state.

The approach rests on three branches of scientific enquiry. The first addresses the scale of human action in relation to the capacity of Earth to sustain it. This is a significant feature of the ecological economics research agenda³¹, drawing on knowledge of the essential role of the life-support properties of the

environment for human wellbeing^{32,33} and the biophysical constraints for the growth of the economy^{34,35}. The second is the work on understanding essential Earth processes^{6,36,37} including human actions^{23,38}, brought together in the fields of global change research and sustainability science³⁹. The third field of enquiry is research into resilience^{40–42} and its links to complex dynamics^{43,44} and self-regulation of living systems^{45,46}, emphasizing thresholds and shifts between states⁸.

Although we present evidence that three boundaries have been overstepped, there remain many gaps in our knowledge. We have tentatively quantified seven boundaries, but some of the figures are merely our first best guesses. Furthermore, because many of the boundaries are linked, exceeding one will have implications for others in ways that we do not as yet completely understand. There is also significant uncertainty over how long it takes to cause dangerous environmental change or to trigger other feedbacks that drastically reduce the ability of the Earth system, or important subsystems, to return to safe levels.

The evidence so far suggests that, as long as the thresholds are not crossed, humanity has the freedom to pursue long-term social and economic development. ■

- Dansgaard, W. *et al.* *Nature* **364**, 218–220 (1993).
- Petit, J. R. *et al.* *Nature* **399**, 429–436 (1999).
- Rioual, P. *et al.* *Nature* **413**, 293–296 (2001).
- Crutzen, P. J. *Nature* **415**, 23 (2002).
- Steffen, W., Crutzen, P. J. & McNeill, J. R. *Ambio* **36**, 614–621 (2007).
- Steffen, W. *et al.* *Global Change and the Earth System: A Planet Under Pressure* (Springer Verlag, 2004).
- Berger, A. & Loutre, M. F. *Science* **297**, 1287–1288 (2002).
- Scheffer, M., Carpenter, S. R., Foley, J. A., Folke C. & Walker, B. H. *Nature* **413**, 591–596 (2001).

Authors

Johan Rockström^{1,2}, Will Steffen^{1,3}, Kevin Noone^{1,4}, Åsa Persson^{1,2}, F. Stuart Chapin, III⁵, Eric F. Lambin⁶, Timothy M. Lenton⁷, Marten Scheffer⁸, Carl Folke^{1,9}, Hans Joachim Schellnhuber^{10,11}, Björn Nykvist¹², Cynthia A. de Wit⁴, Terry Hughes¹², Sander van der Leeuw¹³, Henning Rodhe¹⁴, Sverker Sörlin¹⁵, Peter K. Snyder¹⁶, Robert Costanza¹⁷, Uno Svedin¹, Malin Falkenmark¹⁸, Louise Karlberg¹², Robert W. Corell¹⁹, Victoria J. Fabry²⁰, James Hansen²¹, Brian Walker^{1,22}, Diana Liverman^{23,24}, Katherine Richardson²⁵, Paul Crutzen²⁶, Jonathan A. Foley²⁷

¹Stockholm Resilience Centre, Stockholm University, Kräftriket 2B, 10691 Stockholm, Sweden. ²Stockholm Environment Institute, Kräftriket 2B, 10691 Stockholm, Sweden.

³ANU Climate Change Institute, Australian National University, Canberra ACT 0200, Australia. ⁴Department of Applied Environmental Science, Stockholm University, 10691 Stockholm, Sweden. ⁵Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, Alaska 99775, USA. ⁶Department of Geography, Université Catholique de Louvain, 3 place Pasteur, B-1348 Louvain-la-Neuve, Belgium. ⁷School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK. ⁸Aquatic Ecology and Water Quality Management Group, Wageningen University, PO Box 9101, 6700 HB Wageningen, the Netherlands. ⁹The Beijer Institute of Ecological Economics, Royal Swedish Academy of Sciences, PO Box 50005, 10405 Stockholm, Sweden. ¹⁰Potsdam Institute for Climate Impact Research, PO Box 60 12 03, 14412 Potsdam, Germany. ¹¹Environmental Change Institute and Tyndall Centre, Oxford University, Oxford OX1 3QY, UK. ¹²ARC Centre of Excellence for Coral Reef Studies, James Cook University, Queensland 4811, Australia. ¹³School of Human Evolution & Social Change, Arizona State University, PO Box 872402, Tempe, Arizona 85287-2402, USA. ¹⁴Department of Meteorology, Stockholm University, 10691 Stockholm, Sweden. ¹⁵Division of History of Science and Technology, Royal Institute of Technology, Teknikringen 76, 10044 Stockholm, Sweden. ¹⁶Department of Soil, Water, and Climate, University of Minnesota, 439 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108-6028, USA. ¹⁷Gund Institute for Ecological Economics, University of Vermont, Burlington, VT 05405, USA. ¹⁸Stockholm International Water Institute, Drottninggatan 33, 11151 Stockholm, Sweden. ¹⁹The H. John Heinz III Center for Science, Economics and the Environment, 900 17th Street, NW, Suite 700, Washington DC 20006, USA. ²⁰Department of Biological Sciences, California State University San Marcos, 333 S Twin Oaks Valley Rd, San Marcos, CA 92096-0001, USA. ²¹NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, USA. ²²Commonwealth Scientific and Industrial Organization, Sustainable Ecosystems, Canberra, ACT 2601, Australia. ²³Environmental Change Institute, University of Oxford, Oxford OX1 3QY, UK. ²⁴Institute of the Environment, University of Arizona, Tucson AZ 85721, USA. ²⁵The Faculty for Natural Sciences, Tagensvej 16, 2200 Copenhagen N, Denmark. ²⁶Max Planck Institute for Chemistry, PO Box 30 60, 55020 Mainz, Germany. ²⁷Institute on the Environment, University of Minnesota, 325 VoTech Building, 1954 Buford Avenue, St Paul, MN 55108, USA.

- Lenton, T. M. *et al.* *Proc. Natl Acad. Sci. USA* **105**, 1786–1793 (2008).
- Rockstrom, J. *et al.* *Ecol. Soc.* (in the press); available from http://www.stockholmresilience.org/download/18.1fe8f33123572b59ab800012568/pb_longversion_170909.pdf
- Intergovernmental Panel on Climate Change *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Solomon, S. *et al.*) (Cambridge University Press, 2007).
- Hansen, J. *et al.* *Open Atmos. Sci. J.* **2**, 217–231 (2008).
- Johannessen, O. M. *Atmos. Oceanic Sci. Lett.* **1**, 51–56 (2008).
- Cazenave, A. *Science* **314**, 1250–1252 (2006).
- Church, J. A. & White, N. J. *Geophys. Res. Lett.* **33**, L01602 (2006).
- Mace, G. *et al.* *Biodiversity in Ecosystems and Human Wellbeing: Current State and Trends* (eds Hassan, H., Scholes, R. & Ash, N.) Ch. 4, 79–115 (Island Press, 2005).
- Sala, O. E. *et al.* *Science* **287**, 1770–1776 (2000).
- Sahney, S. & Benton, M. J. *Proc. R. Soc. Lond. B* **275**, 759–765 (2008).
- Diaz, S. *et al.* *Biodiversity regulation of ecosystem services in Ecosystems and Human Well-Being: Current State and Trends* (eds Hassan, H., Scholes, R. & Ash, N.) 297–329 (Island Press, 2005).
- Folke, C. *et al.* *Annu. Rev. Ecol. Evol. Syst.* **35**, 557–581 (2004).
- Chapin, F. S., III *et al.* *Nature* **405**, 234–242 (2000).
- Purvis, A. & Hector, A. *Nature* **405**, 212–219 (2000).
- Foley, J. A. *et al.* *Science* **309**, 570–574 (2005).
- Gruber, N. & Galloway, J. N. *Nature* **451**, 293–296 (2008).
- Machenzie F. T., Ver L. M. & Lerman, A. *Chem. Geol.* **190**, 13–32 (2002).
- Carpenter, S. R. *Regime shifts in lake ecosystems: pattern and variation*, Vol. 15 in *Excellence in Ecology Series* (Ecology Institute, 2003).
- Zillén, L., Conley, D. J., Andren, T., Andren, E. & Björck, S. *Earth Sci. Rev.* **91** (1), 77–92 (2008).
- Bennett, E. M., Carpenter, S. R. & Caraco, N. E. *BioScience* **51**, 227–234 (2001).
- Handoh, I. C. & Lenton, T. M. *Global Biogeochem. Cycles* **17**, 1092 (2003).
- Snyder, P. K., Foley, J. A., Hitchman, M. H. & Delire, C. J. *Geophys. Res. Atmos.* **109**, D21 (2004).
- Costanza, R. *Struct. Change Econ. Dyn.* **2**(2), 335–357 (1991).
- Odum, E. P. *Ecology and Our Endangered Life-Support Systems* (Sinuaer Associates, 1989).
- Vitousek, P. M., Mooney, H. A., Lubchenco, J. & Melillo, J.

M. *Science* **277**, 494–499 (1997).

- Boulding, K. E. The Economics of the Coming Spaceship Earth in *Environmental Quality Issues in a Growing Economy* (ed. Daly, H. E.) (Johns Hopkins University Press, 1966).
- Arrow, K. *et al.* *Science* **268**, 520–521 (1995).
- Bretherton, F. *Earth System Sciences: A Closer View* (Earth System Sciences Committee, NASA, 1988).
- Schellnhuber, H. J. *Nature* **402**, C19–C22 (1999).
- Turner, B. L. II *et al.* (eds) *The Earth as Transformed by Human Action: Global and Regional Changes in the Biosphere over the Past 300 Years* (Cambridge University Press, 1990).
- Clark, W. C. & Dickson, N. M. *Proc. Natl Acad. Sci. USA* **100**, 8059–8061 (2003).
- Holling, C. S. The resilience of terrestrial ecosystems: local surprise and global change in *Sustainable Development of the Biosphere* (eds Clark, W. C. & Munn, R. E.) 292–317 (Cambridge University Press, 1986).
- Walker, B., Holling, C. S., Carpenter, S. R. & Kinzig, A. *Ecol. Soc.* **9**, 5 (2004).
- Folke, C. *Global Environmental Change* **16**, 253–267 (2006).
- Kaufmann, S. A. *Origins of Order* (Oxford University Press, 1993).
- Holland, J. *Hidden Order: How Adaptation Builds Complexity* (Basic Books, 1996).
- Lovelock, J. *Gaia: A New Look at Life on Earth* (Oxford University Press, 1979).
- Levin, S. A. *Fragile Dominion: Complexity and the Commons* (Perseus Books, 1999).

Editor's note This Feature is an edited summary of a longer paper available at the Stockholm Resilience Centre (<http://www.stockholmresilience.org/planetary-boundaries>). To facilitate debate and discussion, we are simultaneously publishing a number of linked Commentaries from independent experts in some of the disciplines covered by the planetary boundaries concept. Please note that this Feature and the Commentaries are not peer-reviewed research. This Feature, the full paper and the expert Commentaries can all be accessed from <http://tinyurl.com/planetboundaries>.

See Editorial, page 447. Join the debate. Visit <http://tinyurl.com/boundariesblog> to discuss this article. For more on the climate, see www.nature.com/climatecrunch.

BOOKS & ARTS

The younger Oppenheimer

Frank Oppenheimer founded the San Francisco Exploratorium: his charisma and passion for science education made him as influential, if not as famous, as his brother, explains **Robert Crease**.

**Something Incredibly Wonderful Happens:
Frank Oppenheimer and the
World He Made Up**

by K. C. Cole

Houghton Mifflin Harcourt: 2009.

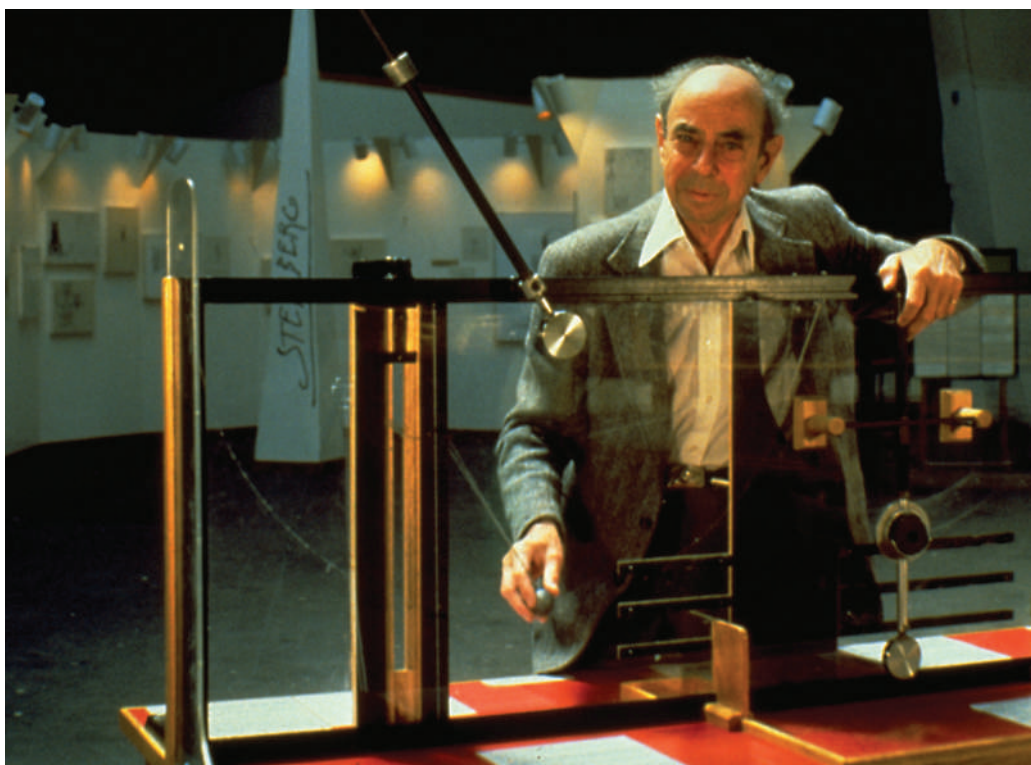
416 pp. \$27

Alfred Russel Wallace wrote that Charles Darwin never lost “the restless curiosity of the child”. One could say the same of the experimental physicist and educator Frank Oppenheimer (1912–1985), younger brother of theoretical physicist J. Robert Oppenheimer, whose life has been far more documented. Like Robert, Frank was involved in leftist politics in ways that damaged his career; unlike Robert, Frank’s relentless enthusiasm allowed him to forge a dramatic comeback. His masterpiece was the San Francisco Exploratorium in California, through which he influenced the lives of countless people.

K. C. Cole, a journalism professor at the University of Southern California in Los Angeles, is one of those people. In the early 1970s, the magazine *Saturday Review* assigned the fledgling writer — who says she had “no interest in science whatsoever” and thought an accelerator was a gas pedal — to cover the Exploratorium. She was transformed by meeting Frank, who struck her as “a kind of Yoda” and helped to launch her career as a science writer. Her enthusiasm is the reason that her book *Something Incredibly Wonderful Happens*, although not deep or probing, is affectionate and evocative.

Frank and Robert were born in Manhattan, New York, to a “little royal family” whose cultivated parents collected art and sent their children to private school. In 1936, while a graduate student in physics at the California Institute of Technology, Frank and his wife Jackie joined the Communist Party; they quit, disenchanted, in 1940. During the Second World War, Frank worked on the Manhattan Project that developed the atomic bomb, watching the blinding explosion of the first nuclear test alongside his brother, who was director of the Los Alamos lab in New Mexico. After the war, Frank studied cosmic rays at the University of Minnesota. In 1949, he was forced to resign and was blacklisted, unable to obtain employment in an academic institution.

Unfortunately, Cole glosses over these



Frank Oppenheimer brought a “rancher’s aesthetic” to the Exploratorium science museum.

years, relying on familiar sources of sometimes doubtful reliability. She does not explore Frank’s membership of the Communist Party, calling it “civic minded” and accepting at face value his remarks that it was a “casual thing”. Oddly, she also glosses over his scientific work. Although she says that Frank found cosmic rays “truly fascinating” and cites an obituary that calls his articles “landmark research”, she doesn’t even tell us the papers’ titles.

Unable to find a job in physics, Frank retired with Jackie in 1949 to an isolated cabin near Pagosa Springs, Colorado, to become a farmer. A “gentle Jewish intellectual from Manhattan”, Frank initially did not even know how to turn grass into hay. He doggedly set out to teach himself, selling an inherited Vincent van Gogh painting to pay his expenses.

Cole is more at home writing about a character in an intimate setting than as an actor on a scientific or political stage. She gives us vivid portraits of Frank delivering newborn animals in the snow, “pulling on the legs of the emerging calf while Jackie read instructions out loud from a veterinary manual”. And she recalls

a neighbour telling her of how Frank once became incensed by a cow’s refusal to enter a pen and began to curse wildly, his expletives echoing back and forth across the valley.

In the 1950s, Frank became a respected community member, an elected representative of local cattlemen and a teacher at the local high school. He galvanized the childrens’ interest in science, making them dissect farm animals and disassemble car engines at the scrapyard. To the bafflement of Colorado officials, students from Pagosa Springs — until then a little-known farming community — began raking in prizes at state science fairs.

In 1959, Frank was offered a teaching job at the University of Colorado, where he thrived. The enthusiastic reception to a talk he gave in 1966 on science and education set him on the track to creating the Exploratorium, which opened in 1969.

Cole’s real subjects are the early years of the Exploratorium and its charismatic, maddening creator. The museum was always a work in progress, having a “rancher’s aesthetic” —

THE EXPLORATORIUM, WWW.EXPLORATORIUM.EDU/E.KUTNICK

exhibits were donated and jerry-built such that a foot-pedal lathe became an electricity generator, or a traffic light became an optics lesson. Events included the dissection of a pig's eye. Children played truant to hang out there, teenagers got high there because of all the perceptual stimulation and adults found it exotic and magical.

Cole writes of her own interactions with Frank. Even simple encounters such as eating a meal could turn into an adventure: he once fashioned a gyroscope for her "out of pats of butter and a bread plate to explain precession". The book strings together many vignettes of Frank in action. It does not matter that his ruminations on life are of uneven value and varying coherence — it captures important parts of the man.

Cole charts the encroaching bureaucratization of the Exploratorium, and the final weeks of Frank's life as he died of cancer. The touching

death scene is disrupted when, to the horror of his wife and children, Frank's mistress arrives to profess her devotion — just one of several eruptions of libidinal chaos.

The vividness of this scene whets our appetite for more insight into the 'royal family' itself. Robert married an alcoholic and was himself an adulterer; his daughter committed suicide. In her book, Cole tells us little of Frank's children, not even when they were born. The Oppenheimer family, we suspect, has yet to reveal all of its dark sides. But by shunning a traditional biographical tapestry, Cole successfully, and at times movingly, limits her focus to Frank's infectious passion for science. ■

Robert Crease is professor and chairman of the Department of Philosophy, Stony Brook University, New York 11794-3750, USA. He is the co-author of *J. Robert Oppenheimer: A Life with Abraham Pais*.
e-mail: rcrease@notes.cc.sunysb.edu

and he and his staff were called cranks.

Given Burton's intent of "poking the establishment in the eye", he might have expected some abuse. But in such cases the best strategy is to stick to one's guns. Alas, Burton is not so thick-skinned. He reveals himself as the insecure country cousin awed by the sophistication of established scientists and their fancy dinner parties. He swayed towards them, minutely measuring the distance between the riverbanks lest he offend the mainstream. Burton tried to replicate the US establishment in Canada, but he was often outbid and exploited by opportunists who used *Perimeter* as a trampoline to boost their US careers.

By the time *Perimeter* matured, five years later, the divide between the quixotic first hires and the new wave was painfully evident. The openness of the early days was replaced by Princeton-style hush-hush and invitation-only meetings. The idealists openly confessed that they wished they could find another benefactor, to "start anew and this time do it right". Something had gone wrong: the sought utopia had become a dystopia.

Scientific originality has become big business: being anti-establishment sounds great. Yet few want to take the risks necessary to achieve it. Originality is encouraged in public pronouncements only to be punished when practical decisions are made. Perhaps *Perimeter's* tale proves that there is no recipe for original science: it happens anarchically and by accident, in spite of, rather than because of, scientific institutions.

Burton had his heart in the right place at the outset and we should have some sympathy for him. And *Perimeter* was a success on many fronts: the building is gorgeous, its bistro is outstanding, the administrative staff enlightened. But the institute has failed to attract good students; its outreach activities have become uninspired; its rebellious streak is now all but gone. Our sympathy for the director dissipates when he fails to shoulder any blame. Add to this the book's purple prose and weak iconoclasm and *First Principles* is bound to irritate as much as disappoint.

Eventually *Perimeter* became a scientific institute just like any other. In 2007, Burton left. ■

João Magueijo is professor of theoretical physics at Imperial College London, Prince Consort Road, London SW7 2BZ, UK. His forthcoming book is *A Brilliant Darkness*, about physicist Ettore Majorana. He was a visiting scientist at *Perimeter* in 2005–07.
e-mail: j.magueijo@imperial.ac.uk

Howard's end at Perimeter

First Principles: The Crazy Business of Doing Serious Science

by Howard Burton

Key Porter Books: 2009. 288 pp.
Can\$24.95

The goal of *First Principles* is good: Howard Burton, founding director of the Perimeter Institute for Theoretical Physics, relates the setting up of the institute in Waterloo, Ontario, Canada, following a donation exceeding Can\$100 million by Mike Lazaridis, creator of the BlackBerry, in 1999. But the book's self-congratulatory tone is a major snag. As reality increasingly conflicts with hype, Burton's account evolves into a sad tale.

The institute's aim was to "make waves, big waves", and it got off to a promising start. Burton — a youthful outsider who had only just finished his physics PhD — went about his job with maverick flair, challenging the scientific establishment, attacking its tribalism and allergy to innovation. Here was an opportunity to do things differently: to promote originality, to flatten hierarchy and empower the young researchers actively driving the field. It sounded utopian, but it was worth a try.

Unfortunately, reality failed to comply with Burton's plan. The best days of this haven of free-thinking came while it was still a 'theoretical' theoretical physics institute — before the scientists arrived. The anecdotes Burton narrates in the chapter 'The Trouble with

Physicists' ring hilariously true. But there was also a fatal flaw in *Perimeter's* concept — scientists tend to define 'originality' as what they personally do. Soon the institute's quest for novelty became hijacked by the agendas of the field's usual culprits, and Burton himself came under attack from them.

"You don't go into the woods to find a lumberjack to run a scientific institute," one senior Canadian physicist jibed, commenting on Burton's past as a 'failed' physicist. "Why did he hire all those losers?" asked another when faced with Burton's first batch of appointments. In some circles *Perimeter* became known as the 'institute of lost causes' in response to Burton's promotion of off-the-beaten-track research,



Howard Burton outside the Perimeter Institute.

A floating island of sustainability

The Waterpod Project

The World's Fair Marina, Flushing, Queens, New York
Until 27 September

In 1970, the artist Robert Smithson proposed his *Floating Island*, a commercial shipping barge full of earth and covered in trees that he wanted to have towed around Manhattan island as if to mock its giddy, man-made skyline. In New York this summer, the artist Mary Mattingly turned the idea into reality on a rented barge almost 300 metres square. Her homage to Smithson is as functional as it is aesthetic; she has designed a post-doomsday survival experiment on which she and a few friends have been living since mid-June. "It's a ship. It's a farm. It's an art residence. It's an installation," Mattingly says.

Mattingly is known for her photographic montages and sculptures that depict nomads roaming the deserts, ice fields or mudflats of an imagined future, carrying their homes on their backs like sea turtles. To put her bleak artistic vision of this future to the test, she decided to give up her apartment and job and build an aquatic self-sustainable structure holding everything you would need to survive a rise in sea levels. Three years in the planning, the Waterpod, as the structure is named, took friends and volunteers little more than a month to construct from recycled and donated materials. With its two geodesic Buckminster Fuller domes, it resembles a floating, do-it-yourself version of the Eden Project in Cornwall, UK.

I visited the ship one sunny evening in August, then moored at Pier 5 in Brooklyn Heights. Walking down the gangplank through a gently rocking allotment garden and past a chicken coop, I entered a six-metre-high domed space. Clad in old advertising hoardings that make its interior skin a colourful collage of outsized diamond watches and other objects of dislocated glamour, it is home to a series of art events.

Lonny Grafman, the Waterpod's 'sustainability adviser', and his students from the environmental-resources engineering department at Humboldt State University, California, designed and built 11 projects to showcase sustainable grassroots technology on the ship. These include a hydroponics system used by Mattingly and her fellow artists, or 'podsters', to grow crops. Other student inventions enable the podsters to collect, purify and recycle



Unlike Manhattan, the self-sustainable Waterpod barge is well-equipped to survive sea-level rises.

water using elaborate catchment and filtration systems; and to rely solely on an off-grid network of solar panels and other renewable energy sources, such as an electricity-generating bicycle. Visitors invariably recoil from the stench of the fly-infested dry-composting toilet.

Mattingly's fellow full-time resident, the artist Alison Ward, took me and my daughter out in a kayak to admire the barge from the water. As we paddled around the stubby, rotting piers of the underused waterfront, my overexcited six-year-old asked if we could stay the night onboard. We soon found ourselves ushered into one of four cabins that looked out on to the Brooklyn Bridge and the Empire State Building beyond.

We picked home-grown vegetables for a candle-lit dinner, which was cooked on a primitive stove and eaten on a long table with spectacular views of the city's lights. The half-dozen guests were offered some of Mattingly's undrinkable home-brewed wine ("it tastes like aspirin" was one verdict). After dinner we watched jellyfish make the river sparkle with phosphorescence.

We awoke early to the sound of ship horns and the heavy rocking caused by water taxis speeding commuters to work. We helped with the morning chores — watering the plants by operating a wooden pump rigged up like a

Stairmaster; feeding the chickens and planting radish seedlings — before cooling off with a swim in the East River and breakfasting on eggs scrambled with the ship's first potatoes. The Waterpod seems more like a floating paradise than a post-apocalyptic ark.

Although Mattingly confesses that there have been occasional bouts of cabin fever, she heads a sanguine crew still thrilled by the kibbutz experience. Admittedly, their utopian experiment hasn't been put through the rigours of endurance and isolation. The fridge is stocked with chocolate and other treats brought by well-wishers. Ward leaves the boat to go to her studio in Manhattan four times a week, and even Mattingly, who tries never to leave the ship, has been tempted off by the prospect of a couple of hot showers.

As the podsters tie everything down for the journey to their next mooring point, there is talk of visiting an animal shelter and adopting a cat. "We should start collecting animals two by two," Mattingly suggests.

Christopher Turner is a writer based in New York. e-mail: drchristopherturner@gmail.com

Correction

The Book Review 'Like minds can be small minds' (*Nature* **461**, 40–41; 2009) incorrectly stated that US voters are now more likely to declare themselves Independents than they were "a century ago". It should read "last century".

M. NAGLE/THE NEW YORK TIMES/EYEVINE

NEWS & VIEWS

L. C. MARIGO/STILL PICTURES; K. R. MORRIS/CORBIS



Ritual centres: the imposing remains of Tikal (left) and Angkor Wat.

ARCHAEOLOGY

Maya, Khmer and Inca

Jared Diamond

Past societies have struggled against environmental problems similar to those that beset us today. Three publications illuminate the outcomes for three different tropical civilizations during the period AD 700–1600.

From time to time, separate archaeological projects on different societies end up by suggesting common themes to events in the ancient world. Thus, two new studies^{1,2} point to parallels between the collapse of cities on opposite sides of the globe — the southern lowland Maya cities in Central America, and Angkor, the centre of the Khmer empire in what is now Cambodia. These parallels include the effects of climate change, which hurt both the Maya and the Khmer. By contrast, as a third report³ indicates, climate change seems to have benefited another ancient civilization, the Incas of South America.

The Maya of Central America, famous for their architecture and writing, abandoned most cities in their southern lowlands between AD 800 and 950 (ref. 4). Lentz and Hockaday¹ have now contributed to our understanding of this collapse by identifying the wood used in 135 support beams from all six major temples and two palaces at the Maya city of Tikal, spanning the century of its peak population (AD 700–810). Although the surrounding forests are rich in tree species, only two species — sapodilla and logwood — accounted for all of the beams. Both share the advantages of being composed of strong, durable wood. Sapodilla is slower growing but superior because it grows much larger; logwood has the drawbacks of thorns, of gnarling with age and of growing in poorly accessible thickets.

Use of these two species shifted with time. All beams sampled from buildings erected early in the eighth century were from sapodilla, often from big trees found only in old-growth forest. During the mid-eighth century, logwood supplanted sapodilla. In the early ninth century, just before temple construction ceased at Tikal, builders resumed using sapodilla, but as smaller beams, presumably from younger trees in secondary forest. These shifts are as expected if builders first used the best and biggest trees, then shifted to suboptimal or smaller trees as preferred trees became depleted.

This study¹ expands on other evidence for deforestation's role in the decline of the southern-lowland Maya. But it seems significant that some old forests with big sapodilla trees were still standing around Tikal in AD 700, nearly two centuries after Tikal's population began to soar. As Lentz and Hockaday note¹, this implies that the Maya practised some forest conservation, perhaps by designating sacred groves or royal forests. Nevertheless, those efforts were ultimately overwhelmed by the demand for timber and farmland.

For the next study², we move halfway around the world, to the Khmer empire's capital of Angkor in modern Cambodia⁵, where a collaboration directed by Roland Fletcher, Michael Barbetti and Daniel Penny has been studying another famous abandonment^{2,6}. Data gathered by NASA radar imaging, ultra-light airplane

overflights and ground surveys have transformed our understanding of Angkor. Early archaeologists could easily recognize Maya and Khmer ritual centres, because Tikal's temples soar to heights of 70 metres, and Angkor Wat is the world's largest religious monument. But the households surrounding these centres did not leave impressive ruins, producing the mistaken view that the centres stood in isolation. The new surveys² show that, at its peak (about AD 1100–1300), Angkor was probably the world's most extensive low-density city, covering 1,000 square kilometres and with half a million or more inhabitants.

Other parallels between Tikal and Angkor include the multifactorial causes of their declines. Both involved population growth and deforestation for timber and agriculture, resulting in erosion and siltation, droughts and fighting. Still another parallel is the slowness of both declines. Whereas Norse Greenland's Western Settlement disappeared in a single winter⁷, the declines at Angkor and the Maya cities stretched over at least a century. But the end result was the same in both the latter cases: populous cities reduced to uninhabited, jungle-covered ruins.

Along with those parallels were big differences between Angkor and the Maya realm. Foremost was the unification of the whole Angkor region into a Khmer empire, whereas the Maya remained splintered into dozens

of warring city states. Was that because the Khmer area's higher agricultural productivity, domestic animals for transport and abundant fish and other protein sources enabled the Khmer — but not the Maya — to control large domains and feed standing armies of conquest? Related to that political difference, the Khmer practised water management on a scale dwarfing that of the Maya and most other regions of the world. Angkor's surrounds were converted into an artificial landscape criss-crossed with canals, embankments, reservoirs, dams and other massive engineering works to redirect river flows, store water for the dry season and avert floods by disposing of excess water during monsoons. The Khmer struggled for centuries to maintain their hydraulic landscape until it became overwhelmed by climate change, producing floods that broke embankments and canals filled with sediments from eroded terrains⁶.

For the third study³, a success story, we return to the New World. Why did the Inca empire of the Andes expand to become the largest Native American empire, only a few centuries after the Wari and Tiwanaku empires of the same region collapsed? Chepstow-Lusty and colleagues³ have analysed a mud core from Lake Maracaoco near the Inca capital of Cuzco, representing 4,200 years of accumulated sediments. By sampling every centimetre over the core's top 1.9 metres, they obtained a temporal resolution of about 6 years. They measured the concentrations of pollen and other plant parts, and of charcoal, and ¹³C/¹²C and C/N ratios, as proxies for local climate, human activity and plant communities.

It turned out that after AD 880 there was increasing drought, which may thus have contributed to the Wari and Tiwanaku collapses, as well as to the earlier and later collapses of the Maya and Khmer. But after AD 1100, during the Northern Hemisphere's Medieval Warm Period, temperatures rose, enabling the Incas to extend agriculture to higher elevations, increase their arable-land area, exploit increased glacial meltwater for irrigation, store more food for their armies, and grow alder trees for nitrogen fixation and timber. Thus, although the Incas' military and administrative organization was essential to their conquests, climate amelioration played a part.

This reminds us that climate can change in either direction, and that in the past such change has variously helped or hurt human societies. But human overexploitation of environmental resources never helps. As Lentz and Hockaday note¹, "Tikal's inhabitants became trapped in a positive feedback loop wherein increasing demands on a shrinking resource base ultimately exceeded the carrying capacity of their immediate environs. The ecological lessons learned from the Late Classic Maya, with their meteoric population increase accompanied by environmental overstretch, serve as a distant mirror for our own cultural trajectory." Amen. ■

Jared Diamond is in the Geography Department, University of California, Los Angeles, California 90095-1524, USA.
e-mail: jdiamond@geog.ucla.edu

1. Lentz, D. L. & Hockaday, B. J. *Archaeol. Sci.* **36**, 1342–1353 (2009).
2. Fletcher, R. *Insights (Durham Univ. Inst. Adv. Study)* **2** (4), 1–19 (2009).

3. Chepstow-Lusty, A. J. *et al. Clim. Past* **5**, 375–388 (2009).
4. Webster, D. *The Fall of the Ancient Maya* (Thames & Hudson, 2002).
5. Coe, M. D. *Angkor and the Khmer Civilization* (Thames & Hudson, 2003).
6. Evans, D. *et al. Proc. Natl Acad. Sci. USA* **104**, 14277–14282 (2007).
7. Panagiotakopulu, E., Skidmore, P. & Buckland, P. *Naturwissenschaften* **94**, 300–306 (2007).

DEVELOPMENTAL BIOLOGY

Rise of the source–sink model

Alexander F. Schier and Daniel Needleman

Gradients of signalling molecules dictate where specific cell types form in developing tissues, but how these gradients are set up is much debated. A model proposed 40 years ago by Francis Crick may provide an answer.

How do the thousands of different cell types in an animal arise time and again at particular locations during embryonic development? The answer lies partly in the distribution of signalling molecules called morphogens¹, which are released from local sources and form concentration gradients in target tissues. Cells that are close to the source of the morphogen are exposed to high signal concentrations and activate developmental programs that differ from those in cells that are farther away and exposed to lower levels of morphogen. This powerful strategy means that the same signalling molecule can be used in the formation of different cell types. But how are morphogen gradients established? On page 533 of this issue, Yu *et al.*² describe one mechanism. They propose that, during the development of the zebrafish embryo, the morphogen fibroblast growth factor 8 (FGF8) spreads rapidly by diffusion from a local source and is then taken up by target tissues. This implies that the combination of free random motion and cellular uptake generates a signalling gradient that endows cells with different developmental fates.

Yu and colleagues' findings² support a model proposed almost 40 years ago by Francis Crick, dubbed the source–sink model. Crick put forward a mechanism³ to explain how morphogen gradients could be set up in a developing tissue. He calculated that a stable gradient can be generated by the local production of a signal at one end of a tissue (the source), its spread into surrounding cells, and its local removal at the other end (the sink). Crick argued as part of the source–sink model that the spreading of the morphogen occurs through Brownian motion — the random thermal motion of molecules — akin to the spreading of a drop of ink in a glass of water. If correct, this would imply that simple diffusion was a plausible mechanism for patterning embryonic tissues. But is there evidence for the source–sink mechanism *in vivo*?

The authors² used fluorescent correlation

spectroscopy (FCS) to analyse the properties of the FGF8 morphogen in zebrafish embryos. FCS is a powerful technology that was introduced in 1972, when it was shown⁴ that measuring fluctuations in fluorescence in a small volume can determine the diffusion properties of labelled molecules in solution. In its modern incarnation, FCS is sufficiently sensitive to probe the dynamics of single molecules. The technique is widely used by biophysicists to measure the behaviours and interactions of proteins, but its use has largely been limited to *in vitro* systems, single-celled organisms and cells in tissue culture.

By contrast, Yu and colleagues² apply FCS to measure the distribution, diffusion and clearance of FGF8 in zebrafish embryos. These embryos are translucent and are therefore ideal for the visualization of the movements of molecules and cells. Yu *et al.* observed that a stable FGF8 gradient forms within 3 hours after production of fluorescently tagged FGF8 in a local region of the early zebrafish embryo. They obtained a diffusion coefficient for FGF8 in the extracellular space of $\sim 50 \mu\text{m}^2 \text{s}^{-1}$, which is strikingly similar to that obtained for molecules of the same size diffusing in water. Therefore, FGF8 seems to move freely and randomly through extracellular space.

But how can such rapidly moving molecules form stable concentration gradients? Yu *et al.* find that extracellular FGF8 has a half-life of only 10–20 minutes. The authors propose that it is the interplay between fast diffusion and the rapid uptake of FGF8 by target-cell endocytosis that creates the gradient. Indeed, green fluorescent protein (GFP), a molecule with similar diffusion properties to FGF8 but much slower clearance, also spreads rapidly but does not form a stable gradient. As predicted by source–sink models, manipulations that increase FGF8 removal, for instance by increasing cellular endocytosis, decrease the range of the FGF8 gradient. By contrast, decreasing endocytosis

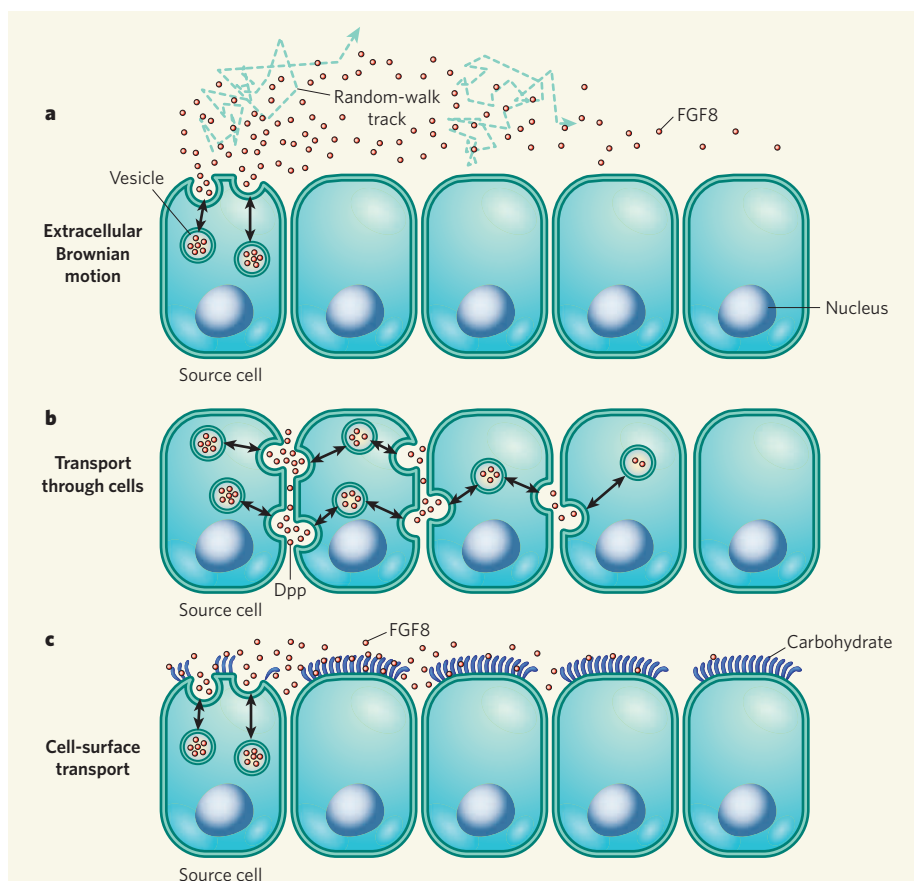


Figure 1 | Models of morphogen dispersal. Source cells harbour vesicles filled with morphogen molecules (red), which fuse with the cell membrane and release their contents. **a**, Yu *et al.*² propose that Brownian motion of molecules in the extracellular space leads to dispersal of the FGF8 morphogen. Two tracks of random walk by single FGF8 molecules are shown. **b**, Kicheva and colleagues⁵ suggest that repeated release and uptake by cells (transcytosis) leads to dispersal of the morphogen Dpp in the fly wing. **c**, A few slowly diffusing FGF8 molecules are associated with carbohydrates at the cell surface³. This cell-surface pool may contribute to long-range dispersal of FGF8.

of FGF8 increases the distance range of the gradient. In this scenario, the sink is not only at one end of the tissue, but is formed by all the target cells that contribute to morphogen removal and gradient formation.

In Yu and colleagues' model², the transport of FGF8 is driven by Brownian motion through extracellular space, as Francis Crick proposed (Fig. 1a). But this might not be the only way of establishing morphogen gradients. In their studies of the morphogen Decapentaplegic (Dpp) protein in *Drosophila*, Kicheva *et al.*⁵ photobleached a large region of the developing wing of the fly and measured the time it took for fluorescently labelled Dpp molecules to re-enter the bleached region. Their result — an effective diffusion coefficient of only $\sim 0.1 \mu\text{m}^2 \text{s}^{-1}$ — means that the typical time taken for Dpp to travel a certain distance would be nearly three orders of magnitude longer than for FGF8. One explanation for this is that the Dpp gradient might be generated by transport through cells⁵ (Fig. 1b). In this case, the repeated cellular uptake and release of Dpp (transcytosis) would lead to an effective diffusion coefficient that is much lower than that of the freely diffusing FGF8.

Why are the results for FGF8 and Dpp so different? The simplest explanation is that these systems really are dissimilar. After all, a fly wing is not a zebrafish embryo, and the principles responsible for establishing

morphogen gradients in these two structures might not be the same.

It is important to keep in mind, however, that different methods were used to measure morphogen transport in the two studies. The FCS experiments provide data on the motion of FGF8 through a volume of $\sim 0.1 \mu\text{m}^3$, whereas the photobleaching experiments measure the recovery of Dpp into a volume of $\sim 10^4 \mu\text{m}^3$, a difference of five orders of magnitude. It might be that the small-scale motion measured by FCS and the large-scale motion observed by photobleaching occur by different mechanisms. If that were the case, FGF8 and Dpp might be dispersed in a similar fashion, and the two studies may have simply probed different aspects of morphogen motion. In this context, it is interesting to note that Yu *et al.*² identified a second, minor fraction of FGF8 that has a 10-fold lower diffusion coefficient. This fraction of FGF8 seems to be associated with carbohydrates located at the surface of cells. The role of this pool of FGF8 is unclear, but it might contribute to the spread of FGF8 along the cell surface (Fig. 1c).

The studies of FGF8² and Dpp⁵ have revealed unexpected complexity of morphogen dispersal. They raise the question of what the biophysical properties of other signalling molecules are during development. Do they function like FGF8 or do they resemble Dpp? The further application of quantitative approaches to other systems and organisms will hopefully address these questions.

Alexander F. Schier and Daniel Needleman are in the Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 0218, USA.
e-mail: schier@mcb.harvard.edu

1. Gurdon, J. B. & Bourillot, P. Y. *Nature* **413**, 797–803 (2001).
2. Yu, S. R. *et al.* *Nature* **461**, 533–536 (2009).
3. Crick, F. *Nature* **225**, 420–422 (1970).
4. Magde, D., Elson, E. & Webb, W. *Phys. Rev. Lett.* **29**, 705–708 (1972).
5. Kicheva, A. *et al.* *Science* **315**, 521–525 (2007).

CLIMATE CHANGE

The El Niño with a difference

Karumuri Ashok and Toshio Yamagata

Patterns of sea-surface warming and cooling in the tropical Pacific seem to be changing, as do the associated atmospheric effects. Increased global warming is implicated in these shifts in El Niño phenomena.

Through the El Niño events^{1,2} that occur every 3–8 years or so, the state of the tropical Pacific Ocean and overlying atmosphere has global effects on climate — sometimes with devastating effects, for example on agriculture in India. El Niños are defined by warmer than normal sea surface temperatures in the eastern tropical Pacific, and are associated with

anomalous atmospheric circulation patterns known as the Southern Oscillation. These coupled phenomena, together called ENSO, have been the subject of research since the late nineteenth century. They remain a matter of intense interest today, not least because of a puzzling shift in behaviour over recent years. That shift, and its possible causes, is the

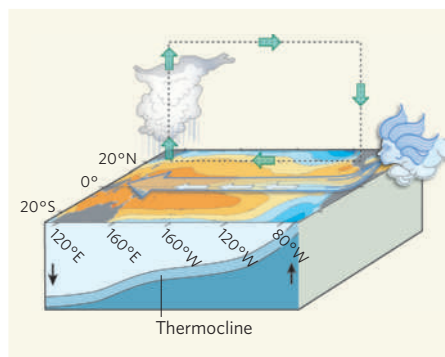


Figure 1 | Normal conditions in the tropical Pacific. Warm surface water and air are pushed to the west by prevailing winds. A consequence is upwelling of cold water on the eastern side, and a shallow thermocline (a subsurface boundary that marks a sharp contrast between warm upper waters and colder deeper waters). Opposite oceanographic conditions prevail on the western side. In the atmosphere, the west is warmer and wetter. Here and in Figure 2, redder colours denote warmer waters, bluer colours denote cooler waters.

subject of the paper by Yeh *et al.*³ that appears on page 511 of this issue.

Normal conditions in the tropical Pacific are shown in Figure 1, and El Niño conditions in Figure 2a. But since the late 1970s, events with increased sea surface temperatures (SSTs) in the central Pacific, sandwiched by anomalous cooling in the east and west, have been observed⁴. These are not like the conventional El Niño: rather, the maximum SST anomaly persists in the central Pacific from the boreal summer through to the winter, modifying the atmospheric circulation and resulting in distinctly different global impacts^{4–10}. This phenomenon has been viewed as a different ‘flavour’ of El Niño⁵, with warming around the Date Line rather than farther east⁶. In other studies^{4,9,10} it has been classified as a new type of tropical Pacific phenomenon, and named El Niño Modoki, or Pseudo El Niño, to stress the differences from the conventional El Niño. Other names, such as Central El Niño¹¹ and Warm Pool El Niño¹², have been proposed. But all of these refer to the same basic phenomenon, shown in Figure 2b, as viewed from slightly different standpoints.

What has caused this recent increased occurrence of the new-flavour of El Niño, with its distinct impact on the northern Pacific SST, compared with the normal form? Some analyses of observational data attribute it to consequences of the prominent global warming⁴ since the late 1970s — which include changes in the zonal tilt of the thermocline^{4,12}, the distinctive subsurface oceanic boundary between warmer and colder water, and weakened low-level easterly winds in the equatorial region. Yeh *et al.*³ have investigated these possibilities further by taking the data set of observed SSTs since the 1850s, and other climate data since the 1950s, and subjecting all this information to detailed analysis in computer models

of ocean–atmosphere behaviour.

Specifically, the authors compare the outcomes of the models used by the Program for Climate Model Diagnosis and Intercomparison (PCMDI). The model outcomes derive from two types of simulation data run through 11 ocean–atmosphere climate models. The control run represents twentieth-century climate change with anthropogenic and natural forcing up to 2000. The other run is from 100 years of a climate-change simulation based on the so-called global-warming SRES A1B scenario, in which levels of carbon dioxide are maintained at about 700 parts per million.

Yeh *et al.* call the canonical El Niño the Eastern Pacific (EP) El Niño, and the new phenomenon the Central Pacific (CP) El Niño. To define the EP events, they use a well-known measure, the NINO3 index, which is SST anomalies averaged over the area 5° N–5° S, 150° W–90° W. To define the CP events, they use a combination of the NINO3 and NINO4 indices, the latter an average of SST anomalies over 5° N–5° S, 160° E–150° W. They analyse the results from the individual model simulations as well as those aggregated by a multi-model, multi-ensemble method, a technique that is applied in operational dynamical seasonal prediction.

Through standardized error analysis, Yeh

et al. find that, in 8 models out of the 11 considered, the occurrence ratio of the CP-El Niño to the EP-El Niño increases in the global-warming scenario as compared to the corresponding twentieth-century climate simulation. The ratio is significantly higher at 95% confidence level in the global-warming simulations of 4 models out of a subset of 6 with more accurate simulations of the currently observed occurrence ratio. Furthermore, the observed remote impacts on SSTs in the Northern Hemisphere are qualitatively evident in the aggregated atmospheric simulations from 9 models. We must take into account that almost all coupled models are biased in representing ocean–atmosphere climatology and the El Niño SST pattern, intensity and frequency. Nonetheless, all these findings support the hypotheses that the increasing frequency of a new type of El Niño — Modoki, CP-El Niño, call it what you will — is due to global warming, and that, as anthropogenic global warming intensifies, we may see more of these events at the expense of the conventional El Niño.

Much more investigation is needed of course. Only four of the models provide statistically significant support, and the coupled models in the PCMDI have flaws in the realistic simulation of basin-wide structure (see Fig. 4 of the authors’ Supplementary

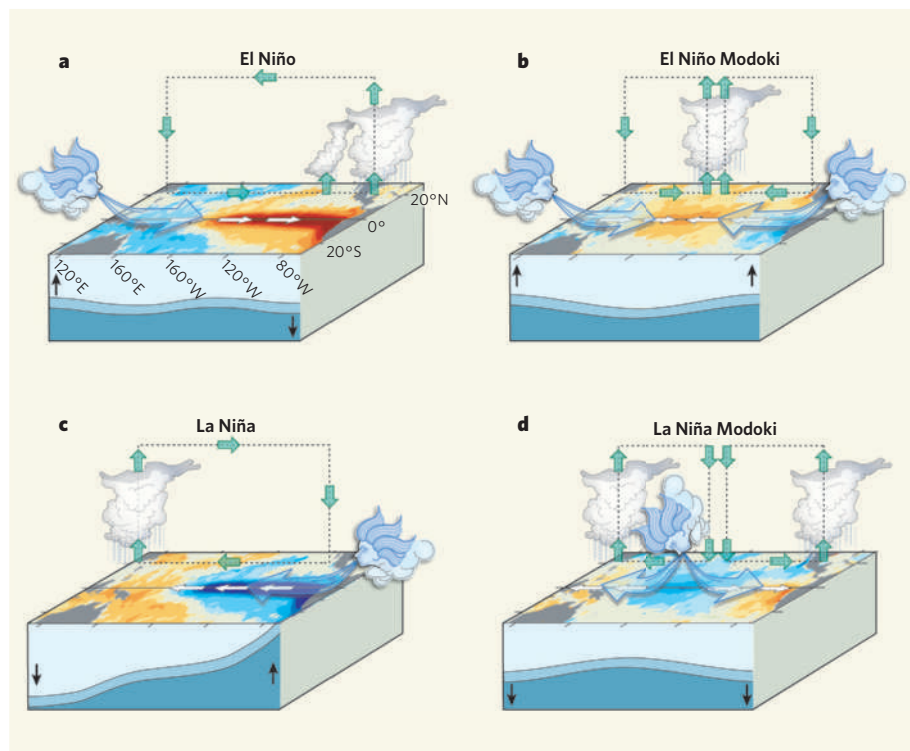


Figure 2 | Anomalous conditions in the tropical Pacific. **a**, An El Niño event is produced when the easterly winds weaken; sometimes, in the west, westerlies prevail. This condition is categorized by warmer than normal sea surface temperatures (SSTs) in the east of the ocean, and is associated with alterations in the thermocline and in the atmospheric circulation that make the east wetter and the west drier. **b**, An El Niño Modoki event is an anomalous condition of a distinctly different kind. The warmest SSTs occur in the central Pacific, flanked by colder waters to the east and west, and are associated with distinct patterns of atmospheric convection. **c**, **d**, The opposite (La Niña) phases of the El Niño and El Niño Modoki respectively. Yeh *et al.*³ argue that the increasing frequency of the Modoki condition is due to anthropogenic warming, and that these events in the central Pacific will occur more frequently if global warming increases.

Information³). Another hypothesis, that the modal shift in El Niño is just a manifestation of natural climate variability on decadal to centennial timescales, also needs to be examined, at least in model simulations. One such study¹³ indicates that both the frequency and the intensity of the ENSO change on multi-centennial timescales.

From the perspective of society's adaptation to climate change, predicting the El Niño Modoki is essential because of its distinctive global impacts on various timescales^{4–11,14,15}. However, our recent experience shows that the effectiveness of the current coupled models in predicting the ENSO Modoki index is low compared with their predictions of the conventional ENSO indices. Further improvement of methods for assimilating observed data into models, as well as improvement of seasonal

prediction models, is necessary. It is also worth examining whether the increasing frequency of the Modoki is related to the recent decadal shift in another ocean–atmosphere phenomenon, the Indian Ocean Dipole: over the past 50 years the SST of the Indian Ocean has increased by about 0.6 °C.

Karumuri Ashok is at the APEC Climate Center, 1463 U-dong, Haeundae-gu, Busan 612-020, Republic of Korea. Toshio Yamagata is at the Application Laboratory, JAMSTEC, 3173-25 Showa-machi, Kanazawa-ku, Yokohama 236-0001, Japan.
e-mails: ashok@apcc21.net (karumuriashok@hotmail.com); yamagata@jamstec.go.jp

1. Rasmusson, E. M. & Carpenter, T. H. *Mon. Weath. Rev.* **110**, 354–384 (1982).

2. Philander, S. G. H. *El Niño, La Niña, and the Southern Oscillation* (Academic, 1990).
3. Yeh, S.-W. *et al. Nature* **461**, 511–514 (2009).
4. Ashok, K., Behera, S. K., Rao, S. A., Weng, H. & Yamagata, T. *J. Geophys. Res.* **112**, doi:10.1029/2006JC003798 (2007).
5. Trenberth, K. E. & Stepaniak, D. P. *J. Clim.* **14**, 1697–1701 (2001).
6. Larkin, N. K. & Harrison, D. E. *Geophys. Res. Lett.* **32**, doi:10.1029/2005GL022738 (2005).
7. Kumar, K. K. *et al. Science* **314**, 115–119 (2006).
8. Wang, G. & Hendon, H. H. *J. Clim.* **20**, 4211–4226 (2007).
9. Weng, H., Ashok, K., Behera, S. K., Rao, S. A. & Yamagata, T. *Clim. Dynam.* **29**, doi:10.1007/s00382-008-0394-6 (2007).
10. Weng, H., Behera, S. K. & Yamagata, T. *Clim. Dynam.* **32**, doi:10.1007/s00382-008-0394-6 (2009).
11. Kao, H.-Y. & Yu, J.-Y. *J. Clim.* **22**, 615–632 (2009).
12. Kug, J.-S., Jin, F.-F. & An, S.-I. *J. Clim.* **22**, 1499–1515 (2009).
13. Wittenberg, A. T. *Geophys. Res. Lett.* **36**, doi:10.1029/2009GL038710 (2009).
14. Kim, H.-M., Webster, P. J. & Curry, J. A. *Science* **325**, 77–80 (2009).
15. Ashok, K., Tam, C.-Y. & Lee, W.-J. *Geophys. Res. Lett.* **36**, doi:10.1029/2009GL038847 (2009).

CHEMICAL BIOLOGY

Caught in the activation

Yi Liu

A crystal structure reveals how a protein kinase is activated by the binding of a small molecule at a pocket far from the catalytic site. This opens the door to the design of modulators of protein phosphorylation.

The protein functions that regulate cell behaviour are themselves tightly regulated by many different inputs. One of these is phosphorylation, the process in which a phosphate group from an ATP molecule is attached to a protein by a protein kinase enzyme¹. There is considerable interest in identifying small molecules that can modulate protein kinase activity, both as research tools and as potential human therapeutic agents. Indeed, eleven kinase inhibitors are already used as drugs for treating cancer, and several more are in clinical trials for the treatment of immune and inflammatory diseases².

Identifying protein kinase inhibitors is relatively easy — it's 'simply' a matter of finding molecules that prevent substrates from binding to the catalytic site of the enzyme. It has proved much more difficult to identify small molecules that activate these enzymes, but a few have been discovered, including those described in 2006 by Engel and colleagues³. Reporting in *Nature Chemical Biology*, the same group (Hindie *et al.*⁴) now disclose the crystal structure of the protein kinase PDK1 in a complex with one of these activators. This is the first crystal structure of a small-molecule activator bound to a protein kinase, and it provides a structural explanation of how this binding activates the enzyme.

A protein kinase catalytic domain consists of a small amino-terminal lobe and a large carboxy-terminal lobe (Fig. 1a). ATP binds in the cleft between the two lobes, whereupon its

terminal phosphate group can be transferred to the protein substrate also bound to the enzyme. Three regions in the ATP-binding site must be precisely positioned for efficient catalysis of the phosphate-transfer reaction: a loop rich in glycine amino acids (the Gly-rich loop); an α -helix

from the small lobe (known as the α -C-helix); and the 'activation' loop from the large lobe. To be activated, most protein kinases must first be phosphorylated in the activation loop⁵, but for full activation some also require a protein to bind to a site remote from the ATP site⁶. In PDK1, for example, a binding site known as the PIF-binding pocket in the small lobe recognizes a phosphorylated amino-acid sequence that activates the kinase⁷.

The authors' PDK1 activators³ bind at the PIF-binding pocket, thus offering the team a golden opportunity to investigate the molecular mechanism of kinase activation mediated by this site. By solving the crystal structure⁴ of a PDK1–activator complex, they found that activator binding causes a change in the conformation

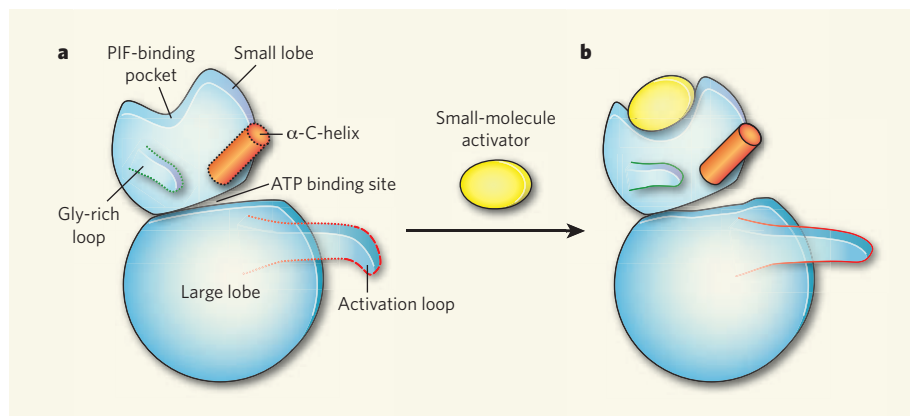


Figure 1 | Small-molecule activation of PDK1. The protein kinase enzyme PDK1 transfers phosphate groups from ATP to a protein substrate, and consists of a large and a small lobe. **a**, Three regions around the ATP-binding site of the enzyme are critical for PDK1 activity — a loop rich in the amino acid glycine (the Gly-rich loop), the α -C-helix and an activation loop. In the inactive state, these regions are flexible and in motion (indicated by dotted lines), and are not aligned for catalysis. Part of the activation loop is unstructured (dashed lines). **b**, Hindie *et al.*⁴ report the crystal structure of PDK1 in a complex with a small-molecule activator of the enzyme. The activator binds at a location known as the PIF-binding pocket, which is remote from the ATP-binding site. Activator binding induces local conformational changes around the PIF-binding pocket, which stabilize the structures of the nearby Gly-rich loop and α -C-helix, leading to conformational changes in these two regions. This stabilization acts as a signal to the disordered region of the activation loop, which adopts an ordered structure in response. Thus, binding of a small-molecule activator to the PIF-binding pocket orients the ATP-binding site into a conformation that is efficient for catalysing phosphate-transfer reactions.

of a key amino-acid residue at the base of the PIF-binding pocket. This conformational change deepens the pocket to better accommodate the small molecule. The activator contains a charged carboxylate group (CO_2^-) that mimics a phosphate group; the interactions of the carboxylate with the surrounding charged and polar amino-acid residues stabilize a conformation of the PIF-binding pocket that is similar to the conformation induced by the binding of a phosphorylated peptide.

When PDK1 is in an inactive conformation, the Gly-rich loop, the activation loop and the α -C-helix are flexible and move around, and part of the activation loop is often disordered. Hindie *et al.*⁴ observed that binding of their activator in the PIF-binding pocket affects all three regions (Fig. 1b). First, local changes in the PIF-binding pocket are transmitted to the nearby Gly-rich loop and to the α -C-helix, causing some residues in these two regions to rearrange. The induced conformational changes in the small lobe then feed through to the activation loop in the large lobe. Finally, the disordered part of the activation loop is ordered and stabilized by interactions with the rearranged Gly-rich loop and α -C-helix. The crystal structure thus provides molecular details of how local changes at the PIF-binding pocket are transduced to the distant ATP-binding site, and how they reorient the ATP site into an active conformation.

The authors found that the conformational changes they observed in the crystal structure of PDK1 upon activator binding also occur in the enzyme in solution. They monitored the structural changes at the ATP site using a fluorescent ATP analogue (TNP-ATP) whose fluorescence intensity increases when it binds to PDK1 at the ATP site. When the authors added an activator to the mix, however, the fluorescence intensity of TNP-ATP decreased, indicating that a conformational change had been produced at the ATP site that prevented TNP-ATP from binding.

In another experiment in solution, Hindie and colleagues⁴ monitored changes at the ATP site using hydrogen–deuterium exchange — an effect in which certain hydrogens in the ‘backbone’ of a protein are replaced with deuterium (D) atoms upon exposure to D_2O . When the authors studied a solution of a fragment of PDK1, they found that less hydrogen–deuterium exchange occurred in the presence of an activator. Apart from the PIF-binding pocket, the sites most protected from exchange included the α -C-helix, the Gly-rich loop and the region just preceding the activation loop. A reduction in hydrogen–deuterium exchange indicates that a conformational change has occurred, so this result tallies well with the changes observed in the crystal structure, further demonstrating that the binding of a small-molecule activator at the PIF-binding pocket induces conformational changes in the ATP site of PDK1 and enhances the activity of the enzyme.

As the *pièce de résistance*, the authors

mutated specific amino acids in PDK1 to identify an amino-acid residue that is important in transducing activator binding at the PIF-binding pocket into the conformational change that activates the enzyme. The crucial residue was identified as a threonine residue located in the activation loop; when the threonine is replaced by a tryptophan, the mutant enzyme still allows small-molecule activators to bind to its PIF-binding pocket, but no activation occurs.

The potency of the PDK1 activators used in this study⁴ is relatively weak, and the carboxylate group present in the molecules may prevent them from entering cells, both of which could prevent these compounds from being developed as drugs. Nevertheless, by providing a molecular basis for protein kinase activation, Hindie and colleagues’ findings are a major advance in the field. They also raise some intriguing questions. For example, could small molecules be designed that induce a different set of conformational changes in the

PIF-binding pocket and inhibit PDK1 activity? Time will tell, but for now this work opens the door to the rational design of small-molecule modulators that target the PIF-binding pocket in PDK1, and in other kinases that have similar regulatory pockets in their small lobes. Such molecules might serve as tools for studying the functions of protein kinases, and as potential new entry points for drug discovery. ■

Yi Liu is in the Department of Research and Development, Intellikine, Inc., 10931 North Torrey Pines Road, La Jolla, California 92037, USA. e-mail: yi@intellikine.com

1. Pawson, T. & Scott, J. D. *Trends Biochem. Sci.* **30**, 286–290 (2005).
2. Zhang, J., Yang, P. L. & Gray, N. S. *Nature Rev. Cancer* **9**, 28–39 (2009).
3. Engel, M. *et al.* *EMBO J.* **25**, 5469–5480 (2006).
4. Hindie, V. *et al.* *Nature Chem. Biol.* **5**, 758–764 (2009).
5. Huse, M. & Kuriyan, J. *Cell* **109**, 275–282 (2002).
6. Pellicena, P. & Kuriyan, J. *Curr. Opin. Struct. Biol.* **16**, 702–709 (2006).
7. Biondi, R. M. *et al.* *EMBO J.* **19**, 979–988 (2000).

ASTROPHYSICS

Inner workings of a star

Sung-Chul Yoon

By borrowing a technique used by seismologists to investigate Earth’s interior, astronomers have probed the hitherto-unknown interior rotation profile of a white-dwarf star.

Just as the sound of a musical instrument is determined by the material from which it is made, so the period with which the brightness of a pulsating star oscillates depends on the specific structure of its interior. Using asteroseismology, astronomers can measure the temporal frequency spectra of pulsating stars — the seismograms of astronomers — to infer information about the stars’ otherwise unobservable interior. For example, rotating pulsating stars ‘sound’ different from non-rotating stars: because rotating stars cannot preserve their spherically symmetric shape during pulsation, their temporal frequency spectrum is marked by non-radial modes of pulsation. This property has allowed astronomers to measure the spin rates of pulsating white dwarfs¹ — stellar remnants of relatively low-mass stars. But probing the internal rotation profiles of these stars is a different matter, and one that astronomers haven’t thus far got to grips with.

On page 501 of this issue, Charpinet *et al.*² present the first evidence that a newly born white dwarf, dubbed PG 1159–035, rotates at the same rate for almost the entire depth of its body. The authors exploited non-radial-pulsation observational data for this star using an innovative method. They constructed a white-dwarf theoretical model that mimics

PG 1159–035 and calculated the resulting pulsation frequencies considering several rotational profiles. By comparing the theoretical model with the data, they concluded that the observed non-radial oscillation frequencies of PG 1159–035 are in fine accord with the interpretation that the star rotates as a solid body. A newly born white dwarf (Fig. 1, overleaf) such as PG 1159–035 is thought to undergo rapid contraction as it loses thermal energy through the emission of neutrinos. Differential rotation, had it been detected by Charpinet and colleagues, would have indicated that the star’s dense inner core and outer envelope were contracting at different rates without rapid transfer of angular momentum between the two.

A typical white dwarf has about 60% of the Sun’s mass, but it is as small as Earth. Assuming that stars maintain their angular momentum throughout their evolution, a reduction in their size as they age would lead to a short spin period — between about 10 seconds and a few minutes — at the white-dwarf stage. But the rotation periods of all single white dwarfs observed to date^{1–3}, including PG 1159–035, are longer than that, having periods of a few days. Together with Charpinet and colleagues’ finding that PG 1159–035 rotates as a solid body, this suggests that single white dwarfs have slow rotation periods because, in the



Figure 1 | Cosmic cocoon. The demise of a dying Sun-like star represents the beginning of a white dwarf. Depicted in this image from the Hubble Space Telescope is an intricate cocoon of gas and dust cast off by one of these stellar remnants as it emerges; the star itself is the white dot (arrowed) in the centre of the image. Charpinet and colleagues' study² indicates that such newly born white dwarfs rotate as solid bodies with relatively long periods of the order of a few days, suggesting that the stellar cores of their progenitors are spun down throughout their evolution.

pre-white-dwarf stages of the stars' evolution, the rotation of their progenitors' cores is slowed down through the efficient transfer and loss of angular momentum from the inner core to the outer parts⁴.

Other evidence supports the idea of the transport of angular momentum in evolving stars. First, the inner region of the Sun, comprising about a fifth to two-thirds of its radius, seems to be rotating as a whole, which would also require a strong angular-momentum transport mechanism⁵. Second, the spin rates of young neutron stars — relics of massive stars that undergo supernova explosions at the end of their life — are much lower than the rates expected if angular momentum were conserved, again implying a rapid slowdown of the rotation of their parent stars' stellar cores during the pre-supernova stages^{6,7}. Therefore, it seems plausible that all stars experience quite rapid transfer of angular momentum and core braking throughout their evolution, leaving slowly rotating remnants at the final stages of their life.

However, we do observe some energetic explosions that can be explained only by exceptionally rapid rotation of their progenitors. These include long-lived γ -ray bursts, which seem to be triggered by a black hole forming in the core of a massive helium star that has an unusually large angular momentum⁸, and some exceptionally luminous type Ia supernovae, which are best explained by the explosion of a white dwarf that exceeds the limiting Chandrasekhar mass by a large amount and that can be explained only by a strong degree of differential rotation^{9,10}. None of these events would have easily occurred had there been efficient transport of angular momentum inside their parent stars, as the above-mentioned studies^{2,4-7} imply. This prompts the question

of why some stars die rotating very rapidly whereas most stars do not.

Addressing that question requires a good understanding of the physical mechanisms responsible for angular-momentum transfer in stellar interiors at various evolutionary stages. This problem is in fact closely related to the more general question of how rotation affects the evolution and subsequent demise of different types of star. In recent studies^{4,7}, theoretical models have been devised that consider a star's magnetic field and the resulting magnetic torque as a possible mechanism for the angular-momentum loss. Although these models are consistent with the spin rates of stellar remnants, notably those of young neutron stars, they might not be able to explain the observed spin rates of single white dwarfs⁴. It is possible that other mechanisms, such as gravity waves, contribute to the angular-momentum loss in low-mass stars⁵.

Future observations of white dwarfs whose traits — for example mass and chemical

composition — differ at different evolutionary stages would greatly improve our understanding of the problem. So far, PG 1159-035 is the only white dwarf to have had its internal rotation determined², but Charpinet and colleagues are already applying their technique to more such stars.

Sung-Chul Yoon is at the Argelander Institute for Astronomy, University of Bonn, Auf dem Hugel 71, D-53121 Bonn, Germany. e-mail: scyoon@astro.uni-bonn.de

1. Fontaine, G. & Brassard, P. *Publ. Astron. Soc. Pacif.* **120**, 1043-1096 (2008).
2. Charpinet, S., Fontaine, G. & Brassard, P. *Nature* **461**, 501-503 (2009).
3. Berger, L. et al. *Astron. Astrophys.* **444**, 565-571 (2005).
4. Suijs, M. P. L. et al. *Astron. Astrophys.* **481**, L87-L90 (2008).
5. Charbonnel, C. & Talon, S. *Science* **309**, 2189-2191 (2005).
6. Ott, C. D. et al. *Astrophys. J. Suppl. Ser.* **164**, 130-155 (2006).
7. Heger, A., Woosley, S. E. & Spruit, H. C. *Astrophys. J.* **626**, 350-363 (2005).
8. Woosley, S. E. *Astrophys. J.* **405**, 273-277 (1993).
9. Howell, D. A. et al. *Nature* **443**, 308-311 (2006).
10. Yoon, S.-C. & Langer, N. *Astron. Astrophys.* **435**, 967-985 (2005).

DEVELOPMENTAL BIOLOGY

A bad boy comes good

Ulrich Theopold

Reactive oxygen species are often blamed for the development of cancer and other diseases. Contrary to their 'bad boy' reputation, these species seem to be essential for the development of immune cells, at least in the fly.

Few concepts have been embraced by popular science as enthusiastically as the idea that reactive oxygen species (ROS) are harmful and that their levels should be controlled by including antioxidants in the diet or as supplements. Despite this 'bad boy' reputation, ROS have important roles in immune defence. They act as potent effector molecules that can kill unwanted intruders, and as intracellular messengers that help to regulate immune responses. On page 537 of this issue, Owusu-Ansah and Banerjee¹ demonstrate an additional function for ROS by showing that they are crucial for the early development of immune cells during the process of blood-cell formation, or haematopoiesis.

The organism providing this insight is the fruitfly *Drosophila melanogaster*, a well-established model system for studying humoral innate immunity — immunity mediated by secreted compounds — but a comparative newcomer to the study of cellular immunity. *Drosophila*'s immune cells come in three classes^{2,3}: plasmotocytes and crystal cells, which are present in uninfected animals, and lamellocytes, which occur mostly in infected animals, where they form capsules around foreign bodies. Plasmotocytes aid wound healing and, like human macrophages and neutrophils, perform a crucial phagocytic role,

ingesting bacteria and dead cells. Alternatively, if infection involves a large number of bacteria, plasmotocytes can form extracellular aggregates, known as nodules, which entrap bacteria. Similar to granulomas in humans, the formation of nodules and capsules is part of local inflammatory reactions. Crystal cells owe their name to the presence of intracellular crystals, which contain an inactive form of phenoloxidase, a key immune-effector enzyme that contributes to the production of substances possessing crosslinking activity that can kill bacteria.

In *Drosophila* larvae, maturation of blood cells occurs in specialized haematopoietic organs, including sub-epidermal sessile blood cells⁴ and the lymph gland, which is located alongside the heart-like dorsal vessel. The Banerjee lab and other groups^{2,3} have found that multipotent blood-cell progenitors are localized to the inner medullary zone of the lymph gland's primary lobe, whereas differentiated cells reside in the outer cortical zone (Fig. 1). The balance between stem cells and differentiated cells depends on the activity of the posterior signalling centre, a specialized haematopoietic niche in the lymph gland^{2,3}.

In their present work¹, the authors' interest was sparked by the observation that ROS are detectable in the medullary zone in progenitor cells that are most akin to the cells that, in

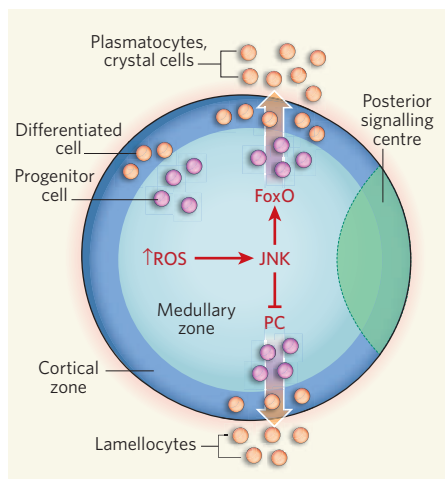


Figure 1 | The good side of reactive oxygen species. Blood-cell development in *Drosophila melanogaster* takes place in the larval lymph glands and is controlled by reactive oxygen species (ROS). Multipotent progenitor cells (pink) in the medullary zone of the primary lobe of the lymph gland differentiate into mature blood cells (orange), which reside in the cortical zone and are subsequently released into the open circulation. Blood-cell differentiation is influenced by the posterior signalling centre, a collection of cells that produce signalling molecules. Owusu-Ansah and Banerjee¹ show that ROS in blood-cell progenitors in the medullary zone are crucial for blood-cell development. Increased amounts of ROS activate the JNK signalling pathway. JNK activation in turn activates the FoxO pathway, leading to differentiation of plasmacytes and crystal cells. Alternatively, repression of Polycomb proteins (PC) by the JNK pathway causes differentiation of lamellocytes. The secondary and tertiary lobes of the lymph gland are not shown in this schematic.

vertebrates, give rise to the myeloid lineages (red blood cells, monocytes and granulocytes). An obvious question is whether there is any connection between the amounts of ROS and haematopoiesis. When Owusu-Ansah and Banerjee¹ decreased the amounts of ROS by inducing the expression of antioxidant scavenger proteins in the lymph gland, the differentiation of progenitor cells to plasmacytes was indeed significantly decreased. Conversely, induction of ROS production led to premature differentiation of all types of *Drosophila* blood cell.

As is seen in other systems, such as mammalian cells⁵, increased amounts of ROS activated a signal-transduction pathway that is often induced by stress, involving an enzyme known as Jun N-terminal kinase (JNK). JNK-pathway activity in turn activated the FoxO pathway, which is known to mediate signalling by insulin and other growth factors. The combined activity of the JNK and FoxO pathways induced differentiation of plasmacytes and crystal cells in the *Drosophila* lymph gland (Fig. 1). Conversely, reduced expression of genes encoding Polycomb proteins downstream of JNK induced the specific production

of lamellocytes (Fig. 1). (Polycomb proteins affect gene expression by modifying the chromatin complex in which DNA is packaged.) Although some of the molecular details of the haematopoietic differentiation pathway in *Drosophila* remain to be elucidated, Owusu-Ansah and Banerjee's findings establish ROS as a major intrinsic effector of blood-cell development.

Evidence that ROS might have a similar role in vertebrates comes from work in adult mice showing that the transcription of genes involved in ROS production is also elevated in myeloid progenitor cells in the bone marrow⁶. This points to an evolutionarily conserved role for ROS during haematopoiesis, although the function of FoxO seems to be less widely conserved⁶, perhaps because the *Drosophila* larval lymph gland represents a developmentally earlier stage than vertebrate bone marrow, corresponding to the fetal liver or other sites of vertebrate embryonic haematopoiesis⁷. The relative contribution of sessile blood cells versus lymph-gland cells to larval or adult immunity in *Drosophila* is also not fully resolved⁴.

So are ROS good or bad? The possibly unsatisfying answer is: it depends. Yes, ROS can do damage if they mix with molecules and cells in the wrong place and at too high a concentration, but they also have many benefits, as highlighted by this study¹. As is often observed in nature, it seems that we must pay an evolutionary price for recruiting potent effector mechanisms to our immune system. The present study comes at a time when the importance of endogenous antioxidant enzymes and the

long-term benefits of the use of antioxidant supplements are increasingly being questioned^{8–10}. In addition, other previously unrecognized functions of ROS, such as a role for hydrogen peroxide in wound healing, have been discovered¹¹. Perhaps the usefulness of antioxidants will have to be judged individually for each chemical and each target, and combinations thereof. For this purpose, studies at the level of the organism are indispensable, as they allow us to trace effects in different organs. Therefore, Owusu-Ansah and Banerjee's study¹ in the fly, and studies in other model organisms such as the nematode worm *Caenorhabditis elegans*^{8,9} and the zebrafish¹¹, should inspire further work on the many roles of ROS and their impact on human health. ■

Ulrich Theopold is in the Department of Molecular Biology and Functional Genomics, Stockholm University, 10691 Stockholm, Sweden. e-mail: uli@molbio.su.se

- Owusu-Ansah, E. & Banerjee, U. *Nature* **461**, 537–541 (2009).
- Crozatier, M. & Meister, M. *Cell. Microbiol.* **9**, 1117–1126 (2007).
- Martinez-Agosto, J. A., Mikkola, H. K., Hartenstein, V. & Banerjee, U. *Genes Dev.* **21**, 3044–3060 (2007).
- Märkus, R. et al. *Proc. Natl Acad. Sci. USA* **106**, 4805–4809 (2009).
- Essers, M. A. et al. *EMBO J.* **23**, 4802–4812 (2004).
- Tothova, Z. et al. *Cell* **128**, 325–339 (2007).
- Cumano, A. & Godin, I. *Annu. Rev. Immunol.* **25**, 745–785 (2007).
- Doonan, R. et al. *Genes Dev.* **22**, 3236–3241 (2008).
- Van Raamsdonk, J. M. & Hekimi, S. *PLoS Genet.* **5**, e1000361 (2009).
- Bjelakovic, G. et al. *J. Am. Med. Assoc.* **297**, 842–857 (2007).
- Niethammer, P., Grabher, C., Look, A. T. & Mitchison, T. J. *Nature* **459**, 996–999 (2009).

HUMAN GENETICS

Tracing India's invisible threads

Aravinda Chakravarti

One measure of the extraordinary level of human diversity found in India is the use of 15 languages on its banknotes. The genetic underpinnings of that population diversity are yielding to whole-genome analysis.

The idea and shape of modern India was an invention of its twentieth-century political leaders, who crafted citizenship defined by civic and universalist, rather than ethnic or religious, criteria precisely because that citizenship is so diverse¹. As Jawaharlal Nehru, the nation's first prime minister, wrote²: “[India] is four hundred million separate individual men and women, each differing from the other ... a bundle of contradictions held together by strong but invisible threads.” Who are these diverse peoples separated by caste, customs and language? Where did they come from, and when? What are the “invisible threads”, beyond claims on the state, that bind them? Studies of biological kinship, which search for the stories of ancestry marked indelibly in a

person's genome, help to provide answers to these questions because they illuminate that unwritten past³. The latest addition to our attempts to understand India through genes comes from Reich, Singh and their colleagues (page 489 of this issue)⁴, who arrive at some bold conclusions about its past population history from genome-variation studies.

The earliest occupation of the subcontinent was by Austro-Asiatic people about 60,000 years ago. They were dispersed and driven into smaller enclaves with the arrival of the Dravidian speakers around 3000 BCE (Before the Common Era, the Common Era marking the same divide as BC and AD). The latter people were themselves driven south with the arrival of the Indo-European speakers in about 1500 BCE.

These early events shaped the growth of an indigenous civilization, with much later conquests by Persians (543 BCE), Alexander III of Macedon (325 BCE), numerous colonial Europeans starting with the Portuguese (1510 CE), and the Mughals (1526 CE). They all came and they were all absorbed — their cultures and their genes — to create the current stew. Although there has been a preoccupation, by both native and foreign scholars, with understanding caste in India and the genetic differences it engenders, there is great diversity at every level: geography, language, caste and customs.

Studies of human variation in India started with the seminal anthropometric surveys of P. C. Mahalanobis in 1941. Subsequently, numerous investigators used various genetic markers (blood groups, serum proteins, enzymes and, later, DNA) to make sense of the vast diversity within the subcontinent. In the genomic era, the Indian Genome Variation Consortium⁵ published a study of 420 single nucleotide polymorphisms (SNPs — base-pair variations in DNA) in 75 genes in 1,871 individuals. The consortium's sample was drawn from 55 groups representing all four language families (Austro-Asiatic, Dravidian, Indo-European, Tibeto-Burman), geography (north, south, east, west), social levels (caste, tribe, religion) and group abundance (small, large), to document the great genomic diversity and the clustering of variation by ethnicity and language (but see ref. 6). The results implied that genetic studies of disease in 'Indians' are hopelessly inadequate unless they account for their specific ancestry. This feature is genetic proof of a population structure first described by the social anthropologist Irawati Karve⁷ as a "patchwork quilt where bits of material of the same colour and shape may be used in a pattern, but where each bit may be of an origin different in place and time".

Reich, Singh and colleagues⁴ instead examine entire genomes' worth of 560,000 SNPs in 132 individuals from 25 groups representing the breadth of social, language and geographic variation in India (see Table 1 and Fig. 1 of their paper on page 490). They sample, in addition, two small groups (the Onge and Great Andamanese) from the Andaman Islands in the Bay of Bengal.

First, the authors show that Indian populations bear the genetic imprint of European, Asian and even, though rarely, African genomes⁵. Second, they find that diversity within India is three to four times greater than that observed within Europe, from which they conclude that many Indian populations, although currently large, were founded by small numbers of individuals with subsequent limited migration⁷. These founder events are dated by the genomic data to between 750 and 2,500 years ago, and therefore occurred well after the arrival of the putative Indo-European speakers.

Third, and most importantly, the authors clearly demonstrate that most of the Indian populations they sampled are mixtures of two

groups that they term ANI (Ancestral North Indians) and ASI (Ancestral South Indians). The degree of ANI:ASI mixture varies between 39% and 71% across India, and is evident in all caste and even tribal groups, and in both extant Indo-European and Dravidian speakers. However, greater ANI ancestry is significantly associated with Indo-European speakers and with traditionally 'higher' caste membership, even after controlling for language. This provides a model of how diversity within India came about. As such, its details are imperfect and will surely be contested, revised and improved; but its implications are significant.

Genetically, the ANI are closest to current-day Europeans whereas the ASI are closest to the disappearing Onge, but neither of these shared ancestries is recent. Reich, Singh *et al.* speculate that the ancestor to both Europeans and ANI spoke a proto-Indo-European language ancestral to both Sanskrit and European languages; the Onge-ASI ancestry is even more remote, and it is unclear whether the ASI were Dravidian speakers. Thus, Indians seem to have a unique set of ancestries for which each population is the same with respect to common descent from two major peoples, but different by virtue of its ancestry proportions and specific genomic content inherited — much like the many hands that can be dealt from a deck of cards. These interpretations are now possible because the authors⁴ have developed new statistical methods to assess specific hypotheses regarding population relationships and ancestry, and also because comparable genomic-variation data on many additional worldwide samples are now available^{8–10}.

The suggestion that each Indian population had small numbers of founders implies strong 'random genetic drift', whereby current frequencies of gene variants depart strongly from their ancestral frequencies simply by chance, thereby increasing genome similarities between members of the same group. This drift effect is largely a result of the early demographic history being shaped by limited numbers of founders, and creates an 'inbreeding' effect whereby genetic variation is lost. This aspect is independent of the additional loss of variation from consanguinity that is found in many parts of India. The cumulative effect is that gene variants may have quite distinct frequencies in India compared with that expected in many other 'related' populations, and that Indians bear the imprint of this very recent local shared ancestry.

This drift and differentiation has four implications. First, studies of relatively few individuals from any Indian population can characterize their common genomic variation adequately. Second, one predicts a high burden of genetically recessive disorders in India, many unique to each population, estimated to

be greater because of local shared ancestry than consanguinity. Third, some diseases will have elevated frequencies in many regions of India owing to shared ANI or ASI ancestry⁵. Fourth, without accounting for local ancestry, genetic association studies can suffer from numerous false positives arising from systematic differences in ancestry between cases and controls. Indeed, language and caste membership may not be adequate control factors.

To a cynic, the existence of the ANI or ASI, their unique and remote ancestry within India, or their suggestive identities as Indo-European and Dravidian speakers, are already common knowledge. But the precise definition of their

ancestral genomic content, their mixture throughout India and the importance of genetic drift are new and have serious implications for both human biology and medicine — and Indian society as well.

Nevertheless, the current analysis⁴ is only a beginning.

The next stage will require samples from a much wider array of populations, including a better sampling of tribal populations and Tibeto-Burman speakers to understand their specific contributions. Indeed, sampling Indians, in the face of their diversity, is a challenge similar to that faced in Africa¹⁰. There is a strong impression that endogamy, the practice of preferring marriage within a group, in India has maintained genetic diversity. However, for this, endogamy must act locally where diverse populations interact⁷: its role can be assessed only by sampling humans locally, not populations distant from one another.

A more comprehensive analysis will require sampling Indians across a grid, assessing both their cultural and genetic diversities, for a deeper understanding of local population structure and the genetic effects of endogamy. Caste and custom may be strong barriers between groups, perhaps even today. But the common shared ancestry and rampant ANI:ASI mixture may be the strong, invisible thread that binds all Indians. ■

Aravinda Chakravarti is at the Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.
e-mail: aravinda@jhmi.edu

"This provides a model of how diversity within India came about. As such, its details are imperfect, but its implications are significant."

1. Khilnani, S. *The Idea of India* (Farrar, Straus & Giroux, 1999).
2. Nehru, J. *The Discovery of India* (Oxford Univ. Press, 1946).
3. Chakravarti, A. *Nature* **457**, 380–381 (2009).
4. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. *Nature* **461**, 489–494 (2009).
5. Indian Genome Variation Consortium *J. Genet.* **87**, 3–20 (2008).
6. Rosenberg, N. A. *et al.* *PLoS Genet.* **2**, e215 (2006).
7. Karve, I. *Hindu Society: An Interpretation* 2nd edn (Deshmukh Prakashan, Poona, 1968).
8. The International HapMap Consortium *Nature* **449**, 851–861 (2007).
9. Li, J. Z. *et al.* *Science* **319**, 1100–1104 (2008).
10. Tishkoff, S. A. *et al.* *Science* **324**, 1035–1044 (2009).

Reconstructing Indian population history

David Reich^{1,2*}, Kumarasamy Thangaraj^{3*}, Nick Patterson^{2*}, Alkes L. Price^{2,4*} & Lalji Singh³

India has been underrepresented in genome-wide surveys of human variation. We analyse 25 diverse groups in India to provide strong evidence for two ancient populations, genetically divergent, that are ancestral to most Indians today. One, the 'Ancestral North Indians' (ANI), is genetically close to Middle Easterners, Central Asians, and Europeans, whereas the other, the 'Ancestral South Indians' (ASI), is as distinct from ANI and East Asians as they are from each other. By introducing methods that can estimate ancestry without accurate ancestral populations, we show that ANI ancestry ranges from 39–71% in most Indian groups, and is higher in traditionally upper caste and Indo-European speakers. Groups with only ASI ancestry may no longer exist in mainland India. However, the indigenous Andaman Islanders are unique in being ASI-related groups without ANI ancestry. Allele frequency differences between groups in India are larger than in Europe, reflecting strong founder effects whose signatures have been maintained for thousands of years owing to endogamy. We therefore predict that there will be an excess of recessive diseases in India, which should be possible to screen and map genetically.

The first systematic surveys of human variation in India focused on anthropometric traits, and found that India is structured along lines of ethnicity as well as geography¹, a result that has since been confirmed by blood group, protein polymorphism^{2,3} and genetic analysis⁴. Genetic studies have further documented differences in relatedness to west Eurasians^{5–8}, and mitochondrial DNA (mtDNA) studies have shown that India contains deep-rooted lineages that share no common ancestry with groups outside of South Asia for tens of thousands of years⁹. The most comprehensive survey of genetic variation in India so far analysed 405 single nucleotide polymorphisms (SNPs) in 55 groups and identified distinct clusters correlated to language and geography¹⁰, while another study analysed 1,200 polymorphisms in 15 Indian American groups¹¹. However, neither study analysed enough data to more finely discern patterns of genetic variation.

We genotyped 132 Indian samples from 25 groups. To survey a wide range of ancestries, we sampled 15 states and six language families (including two language families from the Andaman Islands¹²) (Table 1 and Fig. 1). To compare traditionally 'upper' and 'lower' castes after controlling for geography, we focused on castes from two states: Uttar Pradesh and Andhra Pradesh. We genotyped all samples on an Affymetrix 6.0 array, yielding data for 560,123 autosomal SNPs after filtering (Methods). Allele frequency differentiation between groups was estimated with high accuracy (F_{ST} had an average standard error of ± 0.0011 ; Supplementary Tables 1 and 2). For some analyses, we also merged our data with HapMap¹³ and data from the Human Genome Diversity Panel (HGDP)^{14,15} (Methods).

We analysed these data to address five questions. Does India contain more substructure than Europe? Has endogamy been long-standing in Indian groups? Do nearly all Indians descend from a mixture of populations? Is the ancestry of tribal groups systematically different from castes? What is the origin of the indigenous Andaman Islanders?

Extensive population structure in India

We applied principal components analysis (PCA)^{16,17} to identify outlier groups (Supplementary Fig. 1). The first principal component shows that the Siddi have African ancestry, consistent with their origin involving the Arab slave trade¹⁸. The second shows that the Nyshi and Ao Naga cluster with the Chinese (CHB), consistent with them speaking

Tibeto-Burman languages. The third and fourth show that the Great Andamanese do not cluster tightly, consistent with gene flow from the mainland in the last few generations¹⁹. However, the Onge cluster tightly, making them more useful for studying the relationship of the indigenous Andamanese to groups worldwide (Supplementary Note 1). We treat the Chenchu as a sixth outlier group because of their high minimum F_{ST} of 0.052 from all other groups (Supplementary Table 3).

The average pairwise F_{ST} of the remaining 19 groups is 0.0109. This is much larger than the 0.0033 in a recent study of 23 European groups²⁰, although a strict comparison is difficult, as European studies have focused on cosmopolitan samples^{20,21}, which could underestimate differentiation relative to our village-centred sampling. We considered the possibility that the high F_{ST} could be an artefact due to marriage between close relatives, which is known to be common in southern India²², and which can exaggerate measurements of frequency differentiation. However, when we recalculated F_{ST} correcting for consanguinity²³ (see Appendix in Supplementary Information), the average differentiation decreased only marginally to 0.0100. We also determined that the high F_{ST} was not due to our strategy of sampling diverse groups. Restricting to the nine pairs of groups that were from the same state and traditional caste level, the average inbreeding-corrected F_{ST} was 0.0069; much higher than the analogous 0.0018 in Europe when comparing within regions (Supplementary Table 3).

We propose that the high F_{ST} among Indian groups could be explained if many groups were founded by a few individuals, followed by limited gene flow^{8,24}. This hypothesis predicts that within groups, pairs of individuals will tend to have substantial stretches of the genome in which they share at least one allele at each SNP. We find signals of excess allele sharing in many groups (Supplementary Fig. 2), which as expected tend to occur in the groups that have the highest F_{ST} values from all others ($P = 0.002$ for a correlation). To estimate the age of founder events, we measured the genetic distance scale over which allele-sharing decays, and verified the robustness of our procedure by simulation (Supplementary Fig. 3). Six Indo-European- and Dravidian-speaking groups have evidence of founder events dating to more than 30 generations ago (Supplementary Fig. 2), including the Vysya at more than 100 generations ago (Fig. 2). Strong endogamy must have applied since then (average gene flow less than 1 in 30 per generation) to prevent

¹Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. ³Centre for Cellular and Molecular Biology, Hyderabad 500 007, India. ⁴Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA. *These authors contributed equally to this work.

Table 1 | 25 groups sampled from 13 states of India

Group	Samples	Language family	Traditional caste or social designation	Sampling location				Min F_{ST} to others	
				State/territory	Nearest large town, city or island	Latitude/longitude	Census size†	Uncorrected	Inbreeding corrected
Kashmiri Pandit	5	Indo-European	Upper caste	Kashmir	Dras	34°22'N/75°50'E	7,000	0.0005	0.0023
Vaish	4	Indo-European	Upper caste	Uttar Pradesh	Jaunpur	25°46'N/82°44'E	25,000,000	0.0005	0.0020
Srivastava	2	Indo-European	Upper caste	Uttar Pradesh	Mirzapur	25°10'N/82°37'E	10,000,000	0.0029	0.0023
Sahariya	4	Indo-European	Lower caste	Uttar Pradesh	Allahabad	25°28'N/81°54'E	41,000 ^a	0.0089	0.0087
Lodi	5	Indo-European	Lower caste	Uttar Pradesh	Jhansi	26°45'N/83°24'E	57,000	0.0029	0.0028
Satnami	4	Indo-European	Lower caste	Chhattisgarh	Raipur	20°29'N/85°58'E	4,200,000	0.0038	0.0039
Bhil	7	Indo-European	Tribal	Gujarat	Ahmedabad	23°02'N/72°40'E	7,400,000 ^a	0.0022	0.0027
Tharu	9	Indo-European	Tribal	Uttarkhand	Nainital	29°23'N/79°30'E	96,000 ^a	0.0009	0.0017
Meghawal	5	Indo-European	Lower caste	Rajasthan	Jodhpur	26°18'N/73°04'E	890,000	0.0034	0.0048
Vysya	5	Dravidian	Middle caste	Andhra Pradesh	Anantapur	14°41'N/77°39'E	3,200,000	0.0108	0.0087
Naidu	4	Dravidian	Upper caste	Andhra Pradesh	Chittoor	13°13'N/79°06'E	19,000,000	0.0052	0.0022
Velama	4	Dravidian	Upper caste	Andhra Pradesh	Mahboob Nagar	16°31'N/75°51'E	13,000,000	0.0078	0.0038
Madiga	4	Dravidian	Lower caste	Andhra Pradesh	Warangal	17°58'N/79°35'E	1,600,000 ^b	0.0038	0.0028
Mala	3	Dravidian	Lower caste	Andhra Pradesh	Hyderabad	17°22'N/78°29'E	2,900,000 ^b	0.0038	0.0030
Kamsali	4	Dravidian	Lower caste	Andhra Pradesh	Kurnool	15°49'N/78°02'E	5,100,000	0.0055	0.0022
Chenchu	6	Dravidian	Tribal	Andhra Pradesh	Anantapur	17°22'N/78°28'E	28,000 ^a	0.0524	0.0536
Kurumba	9	Dravidian	Tribal	Kerala	Palakkad	10°54'N/76°27'E	1,300 ^a	0.0021	0.0017
Hallaki	7	Dravidian	Tribal	Karnataka	Uttara Kannada	13°55'N/74°09'E	75,000	0.0072	0.0045
Santhal	7	Austro-Asiatic	Tribal	Jharkhand	Santhal Pargana	24°30'N/87°30'E	2,100,000 ^a	0.0045	0.0057
Kharia	6	Austro-Asiatic	Tribal	Madhya Pradesh	Raigarh	23°08'N/73°07'E	6,900 ^a	0.0045	0.0057
Nyshi	4	Tibeto-Burman	Tribal	Arunachal Pradesh	Papum Pare	26°55'N/92°40'E	56,000 ^a	0.0215	0.0198
Ao Naga	4	Tibeto-Burman	Tribal	Nagaland	Kohima	25°40'N/94°08'E	105,000 ^a	0.0215	0.0198
Siddi	4	Dravidian*	Tribal	Karnataka	Dharwad	15°27'N/75°05'E	25,000	0.0746	0.0757
Onge	9	Jarawa-Onge	Hunter gatherer	Andaman & Nicobar	Little Andaman	10°30'N/92°30'E	97 ^a	0.0905	0.0934
Gr. Andamanese	7	Andamanese	Hunter gatherer	Andaman & Nicobar	Great Andaman	12°12'N/93°00'E	42 ^a	0.0386	0.0414

* The language of the Siddi is Dravidian, but their ancestors spoke a Bantu language.
† Census estimates correspond to all of India. Numbers are based on: ^aref. 50, and ^bref. 51. For some groups (without a superscript) we obtained estimates from the Census of India 1991, Registrar General Office, Government of India.

the genetic signatures of founder events from being erased by gene flow. Some historians have argued that ‘caste’ in modern India is an ‘invention’ of colonialism²⁵ in the sense that it became more rigid under colonial rule²⁶. However, our results indicate that many current distinctions among groups are ancient and that strong endogamy must have shaped marriage patterns in India for thousands of years^{24,27}.

Medical implications

The high frequency differentiation among Indian groups is medically significant as it shows that ‘population stratification’ (systematic

ancestry differences between cases and controls that can lead to false-positive disease associations) may be a confounder in gene-mapping studies. This is superficially at odds with a recent report that in Indian Americans, allele frequency differentiation is lower than among Europeans¹¹. A potential explanation for the discrepancy is that the previous study pooled samples by state of origin, which can mask substructure. For example, when we performed PCA on an independent set of 85 Gujarati Americans²⁸, we found that they separate into two distinct clusters with high differentiation ($F_{ST} = 0.005$) (Supplementary Fig. 4). Similarly, pairs of Uttar Pradesh and Andhra

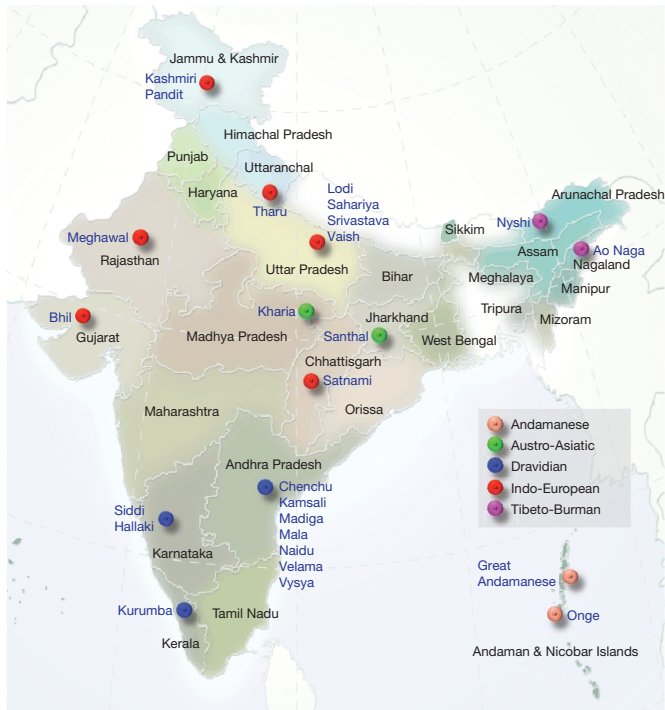


Figure 1 | Map of India. A map of India is shown with the state of origin of the 25 groups that we studied.

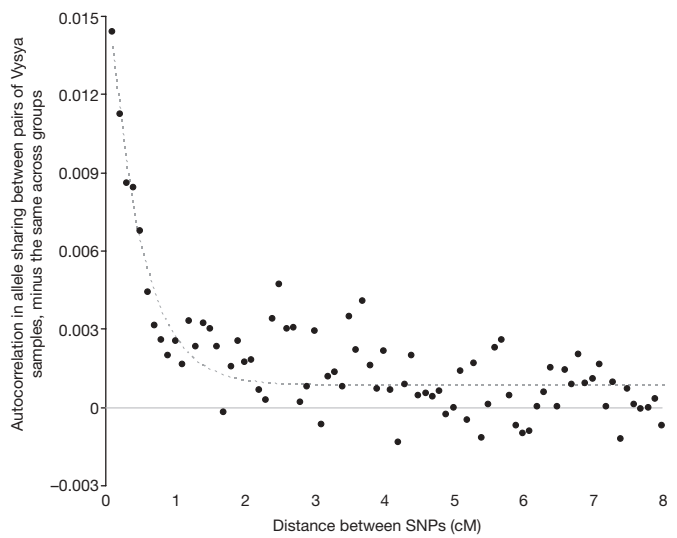


Figure 2 | Linkage disequilibrium based evidence for founder events in India. For each pair of samples, we calculate the autocorrelation of the number of shared alleles as a function of distance, recognizing that SNP genotypes should differ by at most one allele in regions of identity by descent. To correct for background allele sharing, we subtract the same quantity comparing across groups. Allele sharing in the Vysya decreases with an exponential decay of 0.461 cM as shown here, suggesting a founder event roughly $100/(2 \times 0.461) = 108$ generations ago. We present similar analyses for all Indian groups in Supplementary Fig. 2.

Pradesh groups in our data (excluding the outlying Chenchu) have an average F_{ST} of 0.0107, but their differentiation decreases to 0.0033 when we first pool by state. It was recently suggested that to correct for stratification in India, it may be adequate to adjust for membership in five broad genetic clusters¹⁰. However, our results show that many Indian groups have a degree of allele frequency differentiation from their neighbours that is at least as large as that between northern and southern Europeans, which is known to be sufficient to cause false-positive associations to disease if uncorrected²⁹.

The widespread history of founder events in India is also medically significant because it predicts a high rate of recessive disease. In Finland, there is a high rate of recessive diseases that has been shown to be due to a founder event, and that has resulted in a minimum F_{ST} of 0.005 with other European groups²⁰. Our data show that many Indian groups have a minimum F_{ST} with all other groups that is at least as large (Table 1). Haldane wrote decades ago that “if inter-caste marriages in India become common, various... recessive characters will become rarer”³⁰. However, it has not been generally appreciated that this applies to groups throughout India, and not only to groups in which consanguinity is common²². We propose that founder effects are responsible for an even higher burden of recessive diseases in India than consanguinity. To test this hypothesis, we used our data to estimate the probability that two alleles from a group share a common ancestor more recently than that group’s divergence from other Indians, and compared this to the probability that an individual’s two alleles share an ancestor in the last few generations owing to consanguinity²³. Nine of the 15 Indian groups for which we could make this assessment had a higher predicted rate of recessive disease owing to founder events than to consanguinity, including all the Indo-European-speaking groups (Table 2). These results highlight the value of systematically surveying Indian groups to identify those with the strongest founder effects, and prioritizing them for studies to identify recessive diseases and map genes.

A further reason why some diseases are expected to occur at increased frequencies in India is shared descent from a common Indian ancestral population¹⁰. An example is a 25-base-pair deletion in *MYBPC3* that increases heart failure risk by about sevenfold, and occurs at around 4% throughout India but is nearly absent elsewhere³¹. It has recently been shown that the power to discover disease risk

variants can be increased by modelling Indian genetic variation using a reference panel of European and Chinese chromosomes³². However, the example of *MYBPC3* shows that this is an imperfect solution, because clinically significant alleles that are rare outside of India cannot be imputed by studying non-Indian genetic variation. It is important to specifically characterize Indian variation to permit full-powered gene mapping in India.

Population mixture in Indian history

To better understand the genetic ancestry that is only found in India, we carried out a PCA of Europeans (CEU) and Chinese (CHB) along with 22 Indian groups (Fig. 3). The first principal component distinguishes CEU from CHB, and the second reflects ancestry that is unique to India. The most remarkable feature of the PCA is a gradient of proximity to western Eurasians (Supplementary Fig. 5) (an analogous PCA in Europeans did not produce a gradient of proximity to India; Supplementary Fig. 6). We call this the ‘Indian Cline’, and propose that it reflects the fact that different Indian groups have inherited different proportions of ancestry from the ‘Ancestral North Indians’ (ANI) who are related to western Eurasians, and the ‘Ancestral South Indians’ (ASI). To model ANI–ASI mixture, we selected a subset of 18 groups that formed tight clusters along the Indian Cline, and included the Pathan and Sindhi from Pakistan¹⁴ because they were consistent with the Indian Cline in the PCA but showed greater proximity to western Eurasians (Supplementary Note 2), providing more information about ANI–ASI mixture.

To test whether any of the 18 Indian Cline groups were consistent with all ANI or all ASI ancestry, we applied a new 3 Population Test (Methods). If group X is related to groups Y and W by a simple tree (through a history of divergence without subsequent mixture) then if we define the SNP allele frequencies as p_X , p_Y and p_Z , the quantity $(p_X - p_Y)(p_X - p_W)$ averaged over SNPs, should be proportional to the variance in allele frequency since group X split from Y and Z, and thus should be positive. However, this quantity can be negative if X descends from a mixture event (Supplementary Note 3 and Appendix in Supplementary Information). We applied this test to each of the 18 Indian Cline groups in turn using CEU = Y and Santhal = W, and obtained significantly negative scores for 16 groups (Table 2) as assessed by a Block Jackknife analysis that corrects for linkage

Table 2 | Detection and quantification of population mixture along the Indian Cline

Indian Cline group	Samples	Z-score from 3 Population Test for mixture	ANI ancestry (% \pm 1 s.e.)	Genetic drift D from the best fitting combination of ANI and ASI*	Wright’s fixation index F (estimates inbreeding) [†]	Estimated fraction of recessive diseases due to founder events [‡] (%)
Mala	3	−2.5	38.8 \pm 1.2	0.0023	0	100
Madiga	4	−2.7	40.6 \pm 1.2	0.0018	0.0061	23
Chenchu	6	31.3 (n.s.)	40.7 \pm 1.3	0.0492	0	100
Bhil	7	−10.6	42.9 \pm 1.1	0.0024	0	100
Satnami	3	−5.6	43.0 \pm 1.3	0.0019	0	100
Kurumba	6	−12.6	43.2 \pm 1.1	0.0001	0.0052	2
Kamsali	3	−6.5	44.5 \pm 1.3	0.0016	0.0066	19
Vysya	5	5.4 (n.s.)	46.2 \pm 1.2	0.0083	0.0071	54
Lodi	5	−8.9	49.9 \pm 1.1	0.0027	0.0056	32
Naidu	4	−3.3	50.1 \pm 1.2	0.0022	0.0435	5
Tharu	5	−20.6	51.0 \pm 1.2	0.0000	0	NA
Velama	4	−3.2	54.7 \pm 1.3	0.0044	0.0197	18
Srivastava	2	−7.5	56.4 \pm 1.5	0.0023	0	100
Meghawal	5	−13.3	60.3 \pm 1.2	0.0035	0	100
Vaish	4	−22.0	62.6 \pm 1.2	0.0012	0	100
Kashmiri Pandit	5	−20.6	70.6 \pm 1.2	0.0019	0	100
Sindhi	10	−26.3	73.7 \pm 1.1	0.0008	0.0043	16
Pathan	15	−34.3	76.9 \pm 1.1	0.0001	0.0039	3

NA, not applicable; n.s., not significant.

* Estimates of genetic drift (the variance in allele frequencies on any lineage) are based on a model in which each group is a simple mixture of ANI and ASI, followed by subsequent genetic drift specific to that group (corrected for inbreeding). To fit the model, we use the algorithm described in Supplementary Note 4, and fit f_2 , f_3 and f_4 statistics that are calculated in a way that is unbiased by inbreeding (see Appendix in Supplementary Information).

[†] Wright’s fixation index F is estimated as the excess rate at which the two copies of a chromosome within an individual from a group are identical by state, compared within across individuals from that group (see Appendix in Supplementary Information). We set negative values to 0; standard errors are typically \sim 0.003. Owing to the small sample sizes, these estimates are heavily influenced by the samples that happen to have been included in our analysis, and thus should be considered approximate.

[‡] To estimate the proportion of recessive disease cases that are due to founder events, we consider the two alleles that a single individual carries at any locus. With probability F given by Wright’s Fixation Index, they coalesce in the last few generations owing to consanguinity, and with probability $D(1 - F)$, they coalesced since ANI–ASI mixture owing to founder events specific to that group. The fraction of recessive diseases due to founder events can thus be estimated as $D(1 - F)/(F + D(1 - F))$.

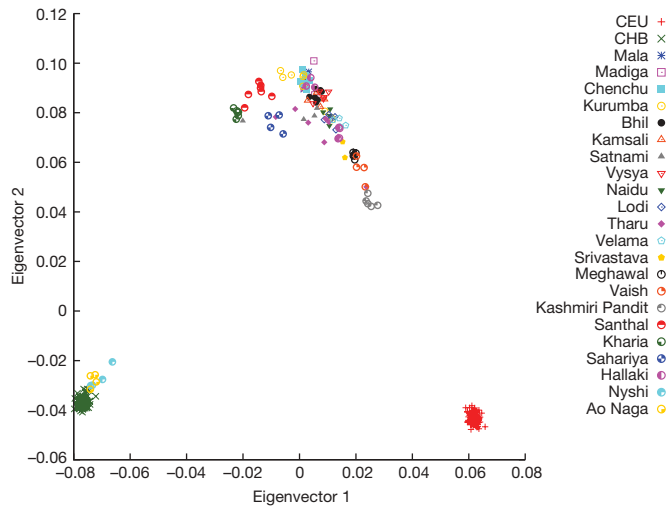


Figure 3 | PCA of 22 groups from the Indian subcontinent. Analysis of these groups along with Europeans (CEU) and Chinese (CHB) shows a gradient of relatedness to CEU that runs through most Indo-European and Dravidian groups, with the Kashmiri Pandit most related to CEU. Both the Austro-Asiatic speaking groups (Kharia and Santhal) and the tribal Sahariya are off-cline, whereas the two Tibeto-Burman speaking groups cluster with CHB. (Data from the outlying Siddi, Onge and Great Andamanese are not shown.)

disequilibrium among SNPs³³ (Methods). These results do not mean that the Indian groups descend from mixtures of European and Austro-Asiatic speakers, but only that they derive from at least two different groups that are (distantly) related to CEU and Santhal.

We verified the evidence of mixture by carrying out a 4 Population Test³⁴. For any four groups there are three possible simple trees. If ((A,B),(C,D)) is correct, the allele frequency differences between A and B should be uncorrelated with those between C and D, which we can assess by averaging the quantity $(p_A - p_B)(p_C - p_D)$ across SNPs (see Appendix in Supplementary Information) and testing for consistency with 0 (Methods). No Indian Cline group could be related simply to CEU, Onge and West Africans (YRI) after testing all trees (Supplementary Table 4).

Relationship of Indians to non-Indians

We developed a model to study the historical relationship of Indian groups to those worldwide, on the basis of the hypothesis that most groups can be approximated as a mixture of two ancestral populations followed by group-specific drift. To fit the model to the data, we computed the squared allele frequency difference between all pairs of groups, and chose parameters by minimizing the difference between observation and expectation (Supplementary Note 4). The idea of fitting allele frequency differentiation to historical models was first explored by Cavalli-Sforza and Edwards³⁵, and here we extend it to trees with mixture. This approach contrasts with the STRUCTURE algorithm, which fits data without a tree³⁶, or a tree in which many groups split simultaneously from an ancestral population followed by mixture³⁷. Although STRUCTURE is accurate for estimating individual mixture proportions in recently mixed groups, it is not clear whether its estimates of ancient mixture are biased because it does not model hierarchical relationships among groups, which could lead to inaccurate estimates of allele frequencies in ancestral populations. In contrast, we use a more realistic tree model, and provide a test of fit.

Applying our model-fitting procedure, we find that the tree (YRI,((CEU,ANI),(ASI, Onge))) provides an excellent fit to the data from Indian groups. In particular, when the Pathan, Vaish, Meghawal and Bhil are modelled as mixtures of ANI and ASI (Fig. 4), the observed allele frequency differentiation statistics are all consistent with the theoretical expectation within three standard deviations (Supplementary Note 4).

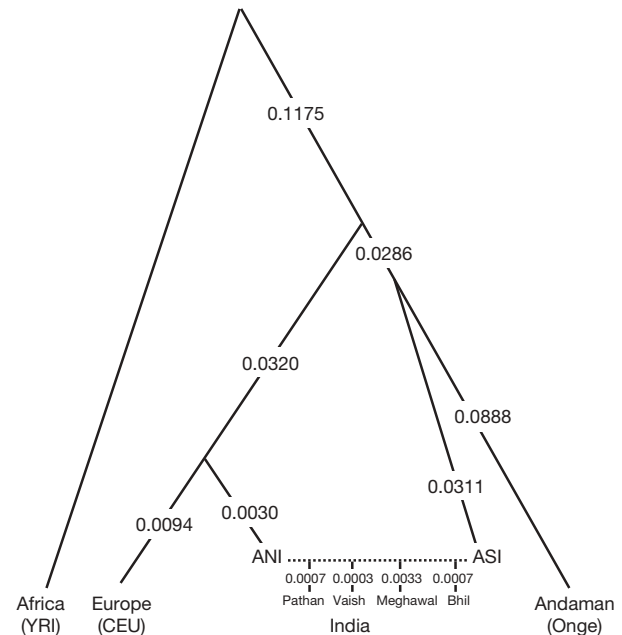


Figure 4 | A model relating the history of Indian and non-Indian groups. Modelling the Pathan, Vaish, Meghawal and Bhil as mixtures of ANI and ASI, and relating them to non-Indians by the phylogenetic tree (YRI,((CEU,ANI),(ASI, Onge))), provides an excellent fit to the data. Although the model is precise about tree topology and ordering of splits, it provides no information about population size changes or the timings of events. We estimate genetic drift on each lineage in the sense of variance in allele frequencies, which we rescale to be comparable to F_{ST} (standard errors are typically ± 0.001 but are not shown).

Two features of the inferred history are of special interest. First, the ANI and CEU form a clade, and further analysis shows that the Adygei, a Caucasian group, are an outgroup (Supplementary Note 4). Many Indian and European groups speak Indo-European languages, whereas the Adygei speak a Northwest Caucasian language. It is tempting to assume that the population ancestral to ANI and CEU spoke 'Proto-Indo-European', which has been reconstructed as ancestral to both Sanskrit and European languages³⁸, although we cannot be certain without a date for ANI–ASI mixture.

Second, our analysis shows that the Onge form a clade with the ASI (Supplementary Note 4), which we verified by running the 4 Population Test on ((YRI,Papuan)(Dai,X)), and finding that it is consistent when X = Onge ($Z = 1.7$) but inconsistent for all Indian Cline groups ($Z \ll -9$) (Supplementary Table 4). Previous mtDNA analyses suggested that the Onge do not share any maternal ancestry with groups outside India within the last ~48,000 years^{19,39}. Although the Onge do share ancestry with some rare haplogroups in some Indian tribal populations within the last ~24,000 years^{39,40}, this observation is consistent with our inferred Onge–ASI clade, as long as the gene flow predated the ASI–ANI mixture that later occurred on the mainland.

We warn that 'models' in population genetics should be treated with caution. Although they provide an important framework for testing historical hypotheses, they are oversimplifications. For example, the true ancestral populations of India were probably not homogeneous as we assume in our model, but instead were probably formed by clusters of related groups that mixed at different times. However, modelling them as homogeneous fits the data and seems to capture meaningful features of history.

Estimates of mixture proportions in India

Estimating the proportions of ANI and ASI ancestry in India is challenging, because we are unaware of any published methods that produce unbiased estimates of mixture proportion in the absence

of accurate ancestral groups. We developed three methods for estimating ancestry, which we verified were accurate even in the face of SNP ascertainment bias and some inaccuracies in our phylogenetic model (Supplementary Note 5), and which we found provided consistent estimates (Supplementary Table 5). The 18 Indian Cline groups all have between 39% and 77% ANI ancestry on the basis of f_3 Ancestry Estimates (Methods), which we quote because it has the smallest standard errors (Table 2). ANI ancestry is significantly higher in Indo-European than Dravidian speakers ($P = 0.013$ by a one-sided test)^{5–8,41}, suggesting that the ancestral ASI may have spoken a Dravidian language before mixing with the ANI⁴². We also find significantly more ANI ancestry in traditionally upper than in lower or middle caste groups ($P = 0.0025$)^{5–8,41}, and find that traditional caste level is significantly correlated to ANI ancestry even after controlling for language ($P = 0.0048$), suggesting a relationship between the history of caste formation in India and ANI–ASI mixture.

We compared our autosomal estimates of ANI ancestry to Y chromosome and mtDNA haplogroup frequencies. Y chromosome analysis has shown that traditionally upper caste and Indo-European speaking groups have increased frequencies of alleles that are also common in western Eurasians^{5,6}. However, mtDNA analysis has shown increased frequencies of haplogroups common in western Eurasians only in northwest India^{7,8,43}. Comparing the autosomal estimates of ANI ancestry to the frequencies of haplogroups characteristic of western Eurasians, we find a significant correlation on the Y chromosome ($P = 0.04$) and a more marginal correlation in mtDNA ($P = 0.08$) (Supplementary Table 6 and Supplementary Fig. 7). The stronger gradient in males, replicating previous reports, could reflect either male gene flow from groups with more ANI relatedness into ones with less, or female gene flow in the reverse direction. The latter hypothesis is unlikely, because extensive female gene flow in India would be expected to homogenize ANI ancestry on the autosomes just as in mtDNA, which we do not observe. Supporting the view of little female ANI ancestry in India, it has been reported that mtDNA ‘haplogroup U’ splits into two deep clades⁴⁴. ‘U2i’ accounts for 77% of copies in India but about 0% in Europe, and ‘U2e’ accounts for 0% of all copies in India but about 10% in Europe. The split is estimated to have occurred about 50,000 years ago, indicating low female gene flow between Europe and India since that time.

Discussion

We have documented a high level of population substructure in India, and have shown that the model of mixture between two ancestral populations, ASI and ANI, provides an excellent description of genetic variation in many Indian groups. A priority for future work should be to estimate a date for the mixture, which may be possible by studying the length of stretches of ANI ancestry in Indian samples^{45,46}, and will shed light on the process leading to the present structure of Indian groups. A second priority should be to discern the details of the history of the ANI and ASI before they mixed, including the date of their separation and their history of expansion and contraction. This may be possible by analysing allele frequency spectrum⁴⁷ and linkage disequilibrium data^{48,49}. Our findings finally have medical implications. By showing that a large proportion of Indian groups descend from strong founder events, these results highlight the importance of identifying recessive diseases in these groups and mapping causal genes.

METHODS SUMMARY

Blood samples were collected with informed consent from volunteers. We designate groups by their anthropological name as well as their geographic location, as it has been shown that both are required to specify an effectively endogamous group in India¹. All DNA samples were genotyped on Affymetrix 6.0 arrays. We restricted most analyses to samples that had no evidence for genetic relatedness and to 560,123 autosomal SNPs for which there were no signs of problematic genotyping and for which the data were relatively complete. For some analyses we also intersected our data with Illumina 650Y genotyping of the Human Genome Diversity Panel¹⁴ and HapMap^{13,28}, which produced a merged data set of 119,744

autosomal SNPs¹⁴. We carried out PCA using the EIGENSOFT software¹⁷, assessed allele frequency differentiation among groups using F_{ST} , assessed inbreeding in each group using Wright’s Fixation Index F^{23} , and computed standard errors using a Block Jackknife³³. To detect the signature of founder events in linkage disequilibrium data, we studied all possible pairs of samples for each group, and recorded whether they share 0, 1 or 2 alleles at each SNP (at SNPs in which both individuals were heterozygous, we recorded 1 allele to be shared to account for the ambiguity in the haplotype phase). Long stretches of allele sharing can reflect regions that are shared identical by descent from a common founder, and by measuring the exponential decay of allele sharing with distance, we inferred the age of the founder event (Supplementary Fig. 3). To test for a history of mixture, we applied 3 and 4 Population Tests (Supplementary Note 3). To infer the proportion of ancestry in each Indian Cline group in the absence of accurate ancestral populations, we used f_3 Ancestry Estimation (Supplementary Note 5).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 April; accepted 5 August 2009.

1. Majumdar, D. N. & Rao, C. R. *Race Elements in Bengal: a Quantitative Study* (Asia Publishing House, 1960).
2. Roychoudhury, A. K. & Nei, M. Genetic relationships between Indians and their neighboring populations. *Hum. Hered.* **35**, 201–206 (1985).
3. Das, B. M., Das, P. B., Das, R., Walter, H. & Danker-Hopfe, H. Anthropological studies in Assam, India. *Anthropol. Anz.* **44**, 239–248 (1986).
4. Zerjal, T. *et al.* Y-chromosomal insights into the genetic impact of the caste system in India. *Hum. Genet.* **121**, 137–144 (2007).
5. Bamshad, M. *et al.* Genetic evidence on the origins of Indian caste populations. *Genome Res.* **11**, 994–1004 (2001).
6. Wells, R. S. *et al.* The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc. Natl Acad. Sci. USA* **98**, 10244–10249 (2001).
7. Thanseem, I. *et al.* Genetic affinities among the lower castes and tribal groups of India: inference from Y chromosome and mitochondrial DNA. *BMC Genet.* **7**, 42 (2006).
8. Basu, A. *et al.* Ethnic India: a genomic view, with special reference to peopling and structure. *Genome Res.* **13**, 2277–2290 (2003).
9. Thangaraj, K. *et al.* In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup ‘M’ in India. *BMC Genomics* **7**, 151 (2006).
10. Indian Genome Variation Consortium. Genetic landscape of the people of India: a canvas for disease gene exploration. *J. Genet.* **87**, 3–20 (2008).
11. Rosenberg, N. A. *et al.* Low levels of genetic divergence across geographically and linguistically diverse populations from India. *PLoS Genet.* **2**, e215 (2006).
12. Abbi, A. Is Great Andamanese genealogically and typologically distinct from Onge and Jarawa? *Language Sciences* doi:10.1016/j.langsci.2008.02.002 (22 April 2008).
13. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
14. Li, J. Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008).
15. Jakobsson, M. *et al.* Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* **451**, 998–1003 (2008).
16. Menozzi, P., Piazza, A. & Cavalli-Sforza, L. Synthetic maps of human gene frequencies in Europeans. *Science* **201**, 786–792 (1978).
17. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
18. Thangaraj, K., Ramana, G. V. & Singh, L. Y-chromosome and mitochondrial DNA polymorphisms in Indian populations. *Electrophoresis* **20**, 1743–1747 (1999).
19. Thangaraj, K. *et al.* Genetic affinities of the Andaman Islanders, a vanishing human population. *Curr. Biol.* **13**, 86–93 (2003).
20. Lao, O. *et al.* Correlation between genetic and geographic structure in Europe. *Curr. Biol.* **18**, 1241–1248 (2008).
21. Novembre, J. *et al.* Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008).
22. Dronamraju, K. R. Mating systems of the Andhra Pradesh people. *Cold Spring Harb. Symp. Quant. Biol.* **29**, 81–84 (1964).
23. Nei, M. & Chesser, R. K. Estimation of fixation indices and gene diversities. *Ann. Hum. Genet.* **47**, 253–259 (1983).
24. Karve, I. *Hindu Society—an Interpretation* (S. R. Deshmukh, 1968).
25. Boivin, N. In *The Evolution and History of Human Populations in South Asia* (eds Petraglia, M. D. & Allchin, B.) 341–362 (Springer, 2007).
26. Dirks, N. B. *Castes of Mind: Colonialism and the Making of Modern India* (Princeton Univ. Press, 2001).
27. Bhasin, M. K. & Walter, H. *Genetics of Castes and Tribes of India* (Kamla-Raj Enterprises, 2001).
28. Index of /genotypes/2008-07_phaseIII. (http://ftp.hapmap.org/genotypes/2008-07_phaseIII/).
29. Campbell, C. D. *et al.* Demonstrating stratification in a European American population. *Nature Genet.* **37**, 868–872 (2005).

30. Haldane, J. B. S. A defense of beanbag genetics. *Perspect. Biol. Med.* **7**, 343–359 (1964).
31. Dhandapany, P. S. *et al.* A common Cardiac Myosin Binding Protein C variant associated with cardiomyopathies in South Asia. *Nature Genet.* **41**, 187–191 (2009).
32. Pemberton, T. J. *et al.* Using population mixtures to optimize the utility of genomic databases: linkage disequilibrium and association study design in India. *Ann. Hum. Genet.* **72**, 535–546 (2008).
33. Künsch, H. R. The jackknife and the bootstrap for general stationary observations. *Ann. Stat.* **17**, 1217–1241 (1989).
34. Keinan, A., Mullikin, J. C., Patterson, N. & Reich, D. Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nature Genet.* **39**, 1251–1255 (2007).
35. Cavalli-Sforza, L. L. & Edwards, A. W. Phylogenetic analysis. Models and estimation procedures. *Am. J. Hum. Genet.* **19**, 233–257 (1967).
36. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
37. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).
38. Mallory, J. P. & Adams, D. O. *The Oxford Introduction to Proto-Indo-European and the Proto-Indo-European World* (Oxford Univ. Press, 2006).
39. Barik, S. S. *et al.* Detailed mtDNA genotypes permit a reassessment of the settlement and population structure of the Andaman Islands. *Am. J. Phys. Anthropol.* **136**, 19–27 (2008).
40. Palanichamy, M. G. *et al.* Comment on “Reconstructing the Origin of Andaman Islanders”. *Science* **311**, 470 (2006).
41. Watkins, W. S. *et al.* Genetic variation in South Indian castes: evidence from Y-chromosome, mitochondrial, and autosomal polymorphisms. *BMC Genet.* **9**, 86 (2008).
42. Southworth, F. C. *Linguistic archaeology of South Asia* (Routledge-Curzon, 2005).
43. Cordaux, R. *et al.* Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur. J. Hum. Genet.* **11**, 253–264 (2003).
44. Kivisild, T. *et al.* Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr. Biol.* **9**, 1331–1334 (1999).
45. Falush, D. *et al.* Traces of human migrations in *Helicobacter pylori* populations. *Science* **299**, 1582–1585 (2003).
46. Baird, S. J. E. Phylogenetics: Fisher’s markers of admixture. *Heredity* **97**, 81–83 (2006).
47. Chikhi, L., Bruford, M. W. & Beaumont, M. A. Estimation of admixture proportions: a likelihood-based approach using Markov chain Monte Carlo. *Genetics* **158**, 1347–1362 (2001).
48. Hellenthal, G., Auton, A. & Falush, D. Inferring human colonization history using a copying model. *PLoS Genet.* **4**, e1000078 (2008).
49. Lohmueller, K. E., Bustamante, C. D. & Clark, A. G. Methods for human demographic inference using haplotype patterns from genomewide single-nucleotide polymorphism data. *Genetics* **182**, 217–231 (2009).
50. Singh, K. S. *People of India, National Series, Volume III, Scheduled Tribes* (Oxford Univ. Press, 1994).
51. Singh, K. S. *People of India, National Series, Volume III, Scheduled Castes* (Oxford Univ. Press, 1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the volunteers from throughout India who donated DNA; A. G. Reddy, A. Shah and R. Tamang for generating the Y chromosome and mtDNA data; J. Neubauer for sample preparation; and A. Tandon for data curation. We thank B. N. Sarkar and A. G. Roy for helping with group census size estimates, and D. Falush, J. Novembre, A. Ruiz-Linares and S. Watkins for comments on the manuscript. D.R., N.P. and A.L.P. were supported by NIH grant HG004168, and D.R. was supported by a Burroughs Wellcome Career Development Award in the Biomedical Sciences. K.T. and L.S. were supported by grants from the Council of Scientific and Industrial Research of the Government of India, and K.T. was supported by a UKIERI Major Award (RG-4772).

Author Contributions K.T. and L.S. collected the DNA samples, D.R., K.T. and L.S. collected the genetic data, N.P. developed the mathematical theory for *f*-statistics, and D.R., K.T., N.P. and A.L.P. analysed the data. D.R. wrote the manuscript and Supplementary Information with input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.R. (reich@genetics.med.harvard.edu) or L.S. (lalji@ccmb.res.in).

METHODS

Sample collection. Blood samples were collected from volunteers with the help of local administrators, and with informed consent and approval of an Institutional Ethical Committee. The names we use are the ones by which the groups are described anthropologically, but are not unique identifiers. We use 'traditionally upper caste' to designate Brahmin and Kshatriya, 'traditionally middle caste' to refer to Vysya, and 'traditionally lower caste' to refer to Shudra. We use 'tribal' and 'hunter gatherer' to refer to non-caste groups.

Genotyping and data curation. We genotyped samples on Affymetrix 6.0 arrays using standard protocols. We restricted analysis to 560,123 SNPs on the autosomes and 27,630 SNPs on the X chromosome with reliable genotyping across >95% of the samples, and used the Birdsuite software⁵² to assign genotypes. We removed ten samples with unusually high relatedness to others as assessed by the rate of genome-wide allele sharing (we included one sample per kinship group). We also intersected our data with HGDP samples genotyped on an Illumina 650Y array¹⁴ and HapMap samples, resulting in 119,744 SNPs on the autosomes and 5,551 SNPs on the X chromosome. As evidence for the usefulness of the merged data set, and the absence of substantial structure in the data related to experimental artefacts, we could not find any PCA that distinguished all the Indians from the HGDP samples.

Statistical methods for analysing population structure. PCA was performed using the EIGENSOFT software¹⁷. We estimated allele frequency differentiation using F_{ST} , which we computed using a formula that has asymptotically minimal variance (see Appendix in Supplementary Information). We also calculated an inbreeding corrected F_{ST} that is asymptotically consistent in the presence of excess homozygosity (see Appendix in Supplementary Information)²³. To compute Wright's Fixation Index F^{23} , an estimate of the inbreeding coefficient for each group, we compared the probability of two alleles being shared identical by state within the same individual, to across individuals from the same group (see Appendix in Supplementary Information).

Block Jackknife procedure to estimate standard errors. To obtain a standard error on F_{ST} as well as the f_2 , f_3 and f_4 statistics, we used a Block Jackknife procedure³³. We divided the genome into contiguous 5 cM chunks and deleted each in turn to quantify the variability of the statistic, which produces a standard error for the value of any estimated quantity. When the null hypothesis indicates that an f -statistic has mean zero as in the 4 Population Test, the jackknife standard error can be converted to a Z-score, which has mean 0 and variance 1 under the null hypothesis. We warn that the normality assumption becomes imperfect for $|Z| > 2$ (not shown). Thus, large Z-scores should be viewed as statistically significant but not simply convertible to P-values⁵³.

Inferring the age of founder events by correlation of allele sharing. For each pair of samples in our data set we record whether they share 0, 1 or 2 alleles at each SNP in the genome. When both individuals are heterozygous we record 1 allele shared (to account for uncertainty about haplotype phase). For each Indian group, we compute the autocorrelation of this allele sharing statistic as a function of distance across all sample pairs, searching for the signature of stretches of allele sharing due to descent from a common founder whose extent reflects the age of the founder event. To correct for background allele sharing inherited from the ancestral populations, we subtract the curve obtained by comparing pairs across groups of similar ANI proportion, choosing from '65 ± 5% ANI' (Meghwal, Vaish and Kashmiri Pandit), '58 ± 5% ANI' (Velama, Srivastava, Meghwal and Vaish), '53% ± 5% ANI' (Lodi, Naidu, Tharu, Velama and Srivastava), '47 ± 5% ANI' (Bhil, Satnami, Kurumba, Kamsali, Vysya, Lodi, Naidu and Tharu) and

'42 ± 5% ANI' (Mala, Madiga, Chenchu, Bhil, Satnami, Kurumba, Kamsali and Vysya). To convert the observed allele sharing decay to a date estimate, we perform a least squares fit to an exponential distribution, $y = a + be^{-2Dt}$. Here, t is the inferred number of generations since the founder event under the assumption of a single strong event, D the genetic distance in Morgans between SNPs, and the factor of 2 reflects the fact that a stretch of allele sharing can be broken by recombination on either haplotype.

3 Population Test for mixture. The 3 Population Test is based on an ' f_3 statistic', a 3-population generalization of F_{ST} . This statistic is equal to the inner product of the frequency differences between a group X and two other groups A and B, which we show in Supplementary Note 3 and the Appendix in the Supplementary Information is proportional to the correlated genetic drift between groups A and X, and groups A and B. If X is related in a simple way (without mixture) to an ancestor, we expect this quantity to be positive, because the genetic drift along the lineage leading from the ancestor to X must be positive. In contrast, if group X has arisen from a mixture of groups related to A and B, it can be negative, and thus the observation of a significantly negative value of the f_3 statistic provides an unambiguous signal of mixture.

4 Population Test for mixture. To assess whether an unrooted phylogenetic tree, for example (YRI,Papuan)(Dai,Onge), is consistent with the SNP allele frequency data, we calculate an ' f_4 statistic', which is expected to be proportional to the correlation in allele frequency differences between pairs of groups (see Appendix in Supplementary Information). If the topology (A,B)(C,D) is correct, then the frequency differences between A and B should reflect genetic drift that is uncorrelated with that between C and D. Thus, the expected value of the product of frequency differences is zero. We compute the statistic $f_4(A;B;C,D)$ with a jackknife standard error. We interpret significant deviations of the f_4 statistics from 0 for all three possible topologies as evidence that the four groups cannot be related by a simple phylogeny without mixture.

f_3 Ancestry Estimation. To obtain estimates of ANI ancestry for each Indian Cline group in the absence of accurate ancestral populations, we used f_3 Ancestry Estimation, f_4 Ancestry Estimation and Regression Ancestry Estimation (Supplementary Note 5), which produce consistent results on the Indian Cline groups as shown in Supplementary Table 5. Here we restrict our description to the f_3 Ancestry Estimates, which we use for Table 2 as this method provides the smallest standard errors. To implement f_3 Ancestry Estimation, we model each Indian Cline group as a linear mixture $K = m_k(\text{ANI}) + (1 - m_k)\text{ASI}$, indicating that each has inherited a proportion m_k of ANI ancestry followed by genetic drift. The topology of Fig. 4 suggests that Onge and ASI are a clade, and hence $f_3(\text{Adygei};\text{Outgroup},K) = m_k f_3(\text{Adygei};\text{Outgroup},\text{ANI}) + (1 - m_k) f_3(\text{Adygei};\text{Outgroup},\text{ASI}) = m_k f_3(\text{Adygei};\text{Outgroup},\text{ANI}) + (1 - m_k) f_3(\text{Adygei};\text{Outgroup},\text{Onge})$. We thus obtain equations: $y_{K,\text{Outgroup}} = (1 - m_k)x_{\text{Outgroup}} + (m_k)z$, where $x_{\text{Outgroup}} = f_3(\text{Adygei};\text{Outgroup},\text{Onge})$ and $y_{K,\text{Outgroup}} = f_3(\text{Adygei};\text{Outgroup},K)$, and solve them using nonlinear least squares, fitting the m_k and z for all three outgroups simultaneously (YRI, Papuan and Dai). We explored whether allowing the coefficient z to depend on x_{Outgroup} improves the fit, as might be expected if the three outgroups do not all have the same position in the phylogeny. We found that this did not change the coefficients m_k or produce a significantly better fit, and hence we allow z to be the same for all three outgroups.

52. McCarroll, S. A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature Genet.* **40**, 1166–1174 (2008).

53. Thorburn, D. On the asymptotic normality of the jackknife. *Scand. J. Stat.* **4**, 113–118 (1977).

A luminal epithelial stem cell that is a cell of origin for prostate cancer

Xi Wang^{1,2,5,6}, Marianna Kruithof-de Julio^{1,2}, Kyriakos D. Economides^{5,7}†, David Walker^{5,6}†, Hailong Yu^{5,6}†, M. Vivienne Halili^{5,6}†, Ya-Ping Hu^{5,6}†, Sandy M. Price^{5,6}, Cory Abate-Shen^{3,4,5,7} & Michael M. Shen^{1,2,5,6}

In epithelial tissues, the lineage relationship between normal progenitor cells and cell type(s) of origin for cancer has been poorly understood. Here we show that a known regulator of prostate epithelial differentiation, the homeobox gene *Nkx3-1*, marks a stem cell population that functions during prostate regeneration. Genetic lineage-marking demonstrates that rare luminal cells that express *Nkx3-1* in the absence of testicular androgens (castration-resistant *Nkx3-1*-expressing cells, CARNs) are bipotential and can self-renew *in vivo*, and single-cell transplantation assays show that CARNs can reconstitute prostate ducts in renal grafts. Functional assays of *Nkx3-1* mutant mice in serial prostate regeneration suggest that *Nkx3-1* is required for stem cell maintenance. Furthermore, targeted deletion of the *Pten* tumour suppressor gene in CARNs results in rapid carcinoma formation after androgen-mediated regeneration. These observations indicate that CARNs represent a new luminal stem cell population that is an efficient target for oncogenic transformation in prostate cancer.

The prostate represents an excellent system for studying the function and molecular regulation of adult epithelial stem cells in the context of both tissue regeneration and cancer. The prostate epithelium is comprised of three differentiated cell types: luminal secretory cells, basal cells and neuroendocrine cells (Fig. 1a)¹. Androgen-deprivation leads to rapid apoptosis of approximately 90% of luminal cells and a small percentage of basal cells, although a stable cell number is maintained in the regressed state^{2,3}. After re-administration of androgens, the prostate epithelium regenerates over roughly 2 weeks²⁻⁴, and is capable of more than 15 rounds of serial regression/regeneration^{5,6}, indicating that the prostate epithelium contains a long-term population of castration-resistant stem cells.

Substantial evidence supports the existence of a basal stem cell population in the prostate⁷, consistent with analyses of progenitor cells in other epithelial tissues⁸. In particular, subpopulations of basal cells isolated using cell-surface markers show bipotentiality and self-renewal in explant culture and tissue grafts⁹⁻¹³. Furthermore, single Lin⁻ Sca-1⁺ CD133⁺ CD44⁺ CD117⁺ cells, which are predominantly basal in the mouse and exclusively basal in the human, can reconstitute prostatic ducts in renal grafts¹⁴. However, explants from *p63* (also known as *Trp63*)-null mice can form prostate tissue and undergo several rounds of serial regression/regeneration in the absence of basal cells¹⁵, suggesting the existence of a distinct luminal stem cell population. Until now, however, luminal stem cells have not been identified in the prostate or other stratified epithelial tissues.

Although basal stem/progenitor cells have been proposed to represent a cell type of origin^{7,16,17}, human prostate cancer has a markedly luminal phenotype. Notably, the absence of basal cells is a diagnostic feature for prostate adenocarcinoma^{18,19}, suggesting either that prostate cancer arises from a luminal cell, or that oncogenic transformation of a basal progenitor results in rapid differentiation of luminal progeny. Here we show that expression of the *Nkx3-1* homeobox gene in the

androgen-deprived prostate epithelium marks a rare luminal cell population that displays stem/progenitor properties during prostate regeneration. Our findings also indicate the relevance of this luminal stem cell population as a cell type of origin for prostate cancer.

Detection of CARNs in the prostate

The *Nkx3-1* homeobox gene regulates prostate epithelial differentiation, and is frequently inactivated at early stages of prostate tumorigenesis²⁰. Notably, *Nkx3-1* homozygous mutant mice develop prostatic intra-epithelial neoplasia (PIN), a precursor of prostate cancer, by 1 year of age²¹⁻²³. In the intact adult mouse prostate, all luminal cells express *Nkx3-1*, and 9.5% of p63⁺ basal cells ($n = 4,291$) also express *Nkx3-1* (Fig. 1b and Supplementary Fig. 1a)²⁴. Previous studies have shown that *Nkx3-1* expression in prostate epithelial cells is reduced or abolished in the absence of androgens *in vivo*, and is consequently androgen-dependent^{25,26}. Thus, *Nkx3-1* expression is rapidly lost after castration, and *Nkx3-1* expression is quickly restored after androgen re-administration to induce prostate regeneration (Fig. 1c, d and Supplementary Fig. 1b).

However, *Nkx3-1* expression is not completely absent in the regressed prostate, but is instead retained in a rare population of epithelial cells (Fig. 1c, e). These castration-resistant *Nkx3-1*-expressing cells (CARNs) comprise 0.7% of total epithelial cells ($n = 38,329$) in the anterior prostate of androgen-deprived males, or approximately 460 CARNs per mouse (Supplementary Table 1, Methods). Furthermore, CARNs are frequently clustered (Fig. 1e), and can be detected in the ventral and dorsal prostate, as well as after a second round of regression (Supplementary Fig. 1c-f).

Notably, all CARNs in the regressed prostate are strictly luminal, because they never express the basal cell marker p63 ($n = 0$ out of 379) or the neuroendocrine marker synaptophysin ($n = 0$ out of 610) (Fig. 1f and Supplementary Table 1). Instead, CARNs express the

¹Department of Medicine, ²Department of Genetics and Development, ³Department of Urology, and ⁴Department of Pathology and Cell Biology, Herbert Irving Comprehensive Cancer Center, Columbia University College of Physicians and Surgeons, New York, New York 10032, USA. ⁵Center for Advanced Biotechnology and Medicine, ⁶Department of Pediatrics, and ⁷Department of Medicine, UMDNJ-Robert Wood Johnson Medical School, Piscataway, New Jersey 08854, USA. †Present addresses: Department of Biological Sciences, Sanofi-Aventis, Bridgewater, New Jersey 08807, USA (K.D.E.); Department of Molecular Biology, Bristol-Myers Squibb Research Institute, Princeton, New Jersey 08543, USA (D.W.); Department of Food Science, Rutgers University, Piscataway, New Jersey 08901, USA (H.Y.); Cardiovascular Diseases Group, Merck Research Laboratories, Rahway, New Jersey 07065, USA (M.V.H.); Johnson and Johnson Skin Research Center, Skillman, New Jersey 08558, USA (Y.-P.H.); Department of Medical Oncology, Cancer Institute of New Jersey, New Brunswick, New Jersey 08903, USA (S.M.P.).

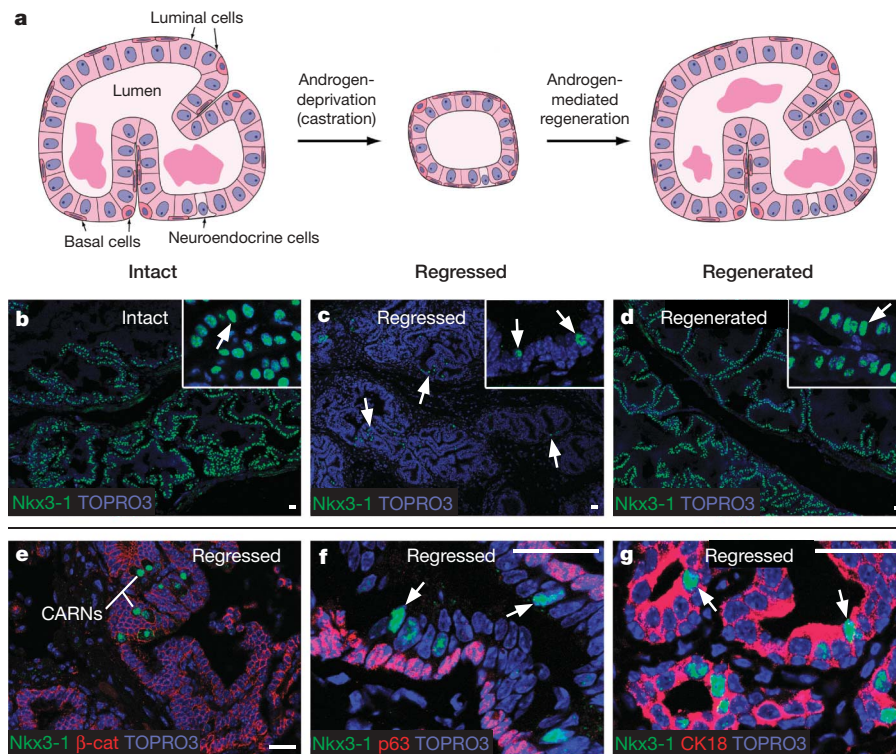


Figure 1 | Expression of *Nkx3-1* in epithelial cells of the intact, regressed and regenerated anterior prostate. **a**, Schematic prostate duct in the intact, regressed and regenerated states. Most luminal cells undergo apoptosis during regression, whereas most basal cells survive; hence, the process of regeneration primarily produces luminal cells. **b**, *Nkx3-1* expression in all luminal cells of the wild-type intact prostate. **c**, *Nkx3-1* expression is mostly

luminal markers cytokeratin 18 (CK18, also known as Krt18) ($n = 828$ out of 837) and androgen receptor ($n = 46$ out of 46), and are growth-quiescent, as they do not co-express Ki67 (also known as Mki67; $n = 0$ out of 151) (Fig. 1g and Supplementary Fig. 1g, h). The CARN population is non-overlapping with the $\text{Lin}^- \text{Sca-1}^+ \text{CD133}^+ \text{CD44}^+ \text{CD117}^+$ stem/progenitor population¹⁴, because CD117 (also known as Kit)-positive cells in regressed prostate are never luminal ($n = 0$ out of 79) (Supplementary Fig. 1k, l). Furthermore, because the CARN population is strictly luminal, it is also distinct from other previously described prostate stem cell populations that are exclusively basal^{9,10,13}.

Bipotentiality and self-renewal

To investigate whether the CARN population might correspond to prostate epithelial progenitors, we performed *in vivo* lineage-marking using a knock-in allele that places a tamoxifen-inducible Cre recombinase^{27,28} under the transcriptional control of the *Nkx3-1* promoter (Supplementary Fig. 2a). We assessed the specificity of this tamoxifen-inducible *Nkx3-1*^{CreERT2} allele in control crosses with the *R26R-YFP* Cre-reporter²⁹ and the *R26R-lacZ* alleles³⁰, and found that Cre-mediated recombination after tamoxifen administration closely recapitulates the endogenous pattern of *Nkx3-1* expression in the intact prostate (Supplementary Fig. 2b–f).

We performed lineage-marking of CARNs by tamoxifen treatment of castrated *Nkx3-1*^{CreERT2/+}; *R26R-YFP/+* or *Nkx3-1*^{CreERT2/+}; *R26R-lacZ/+* adult males (Fig. 2a, b). As expected for genetic marking of CARNs in regressed prostate, we observed yellow fluorescent protein (YFP) or β -galactosidase expression in rare epithelial cells that were strictly luminal (Supplementary Table 1). These lineage-marked cells were never positive for the basal markers p63 ($n = 0$ out of 98) or CK14 (also known as Krt14; $n = 0$ out of 131), and almost never positive for CK5 (Krt5; $n = 2$ out of 93), but always expressed the luminal markers CK18 ($n = 123$ out of 123) and androgen receptor

absent in regressed prostate, except for rare castration-resistant *Nkx3-1*-expressing cells (CARNs, arrows). **d**, Expression of *Nkx3-1* in regenerated prostate, showing similarity to **b**. **e**, Immunostaining for *Nkx3-1* and β -catenin shows clustering of CARNs. **f**, **g**, CARNs are strictly luminal, as shown by the lack of co-staining for *Nkx3-1* (arrows) and p63 (**f**), and by co-localization of *Nkx3-1* (arrows) with CK18 (**g**). Scale bars, 25 μm .

($n = 94$ out of 94) (Fig. 2c and Supplementary Fig. 3a–d). After regeneration, the percentage of lineage-marked cells increased ninefold (from 0.37% ($n = 19,825$) to 3.3% ($n = 95,017$), $P < 0.0001$) (Fig. 2d), indicating the proliferative potential of CARNs. Although most of the lineage-marked cells in regenerated prostates were luminal, we observed occasional $\text{YFP}^+ \text{CK5}^+$, $\text{YFP}^+ \text{p63}^+$, or $\beta\text{-gal}^+ \text{CK14}^+$ basal cells, corresponding to 3.0% of lineage-marked cells ($n = 559$) (Fig. 2e and Supplementary Fig. 3g–i); this percentage of regenerated basal cells is consistent with the low percentage of basal cells lost during regression². Because all of the lineage-marked cells were luminal in the regressed prostate, but could give rise to both basal and luminal cells during regeneration, we conclude that the initial CARN population contains bipotential progenitors.

To investigate the self-renewal of CARNs, we examined whether they could undergo at least one cell division during prostate regeneration to generate a daughter cell that is also a CARN. We determined whether lineage-marked CARNs in castrated *Nkx3-1*^{CreERT2/+}; *R26R-YFP/+* mice would incorporate BrdU during prostate regeneration, while retaining CARN identity (*Nkx3-1* expression) after a subsequent prostate regression (Fig. 2f and Supplementary Fig. 3m). Such triple-positive $\text{Nkx3-1}^+ \text{YFP}^+ \text{BrdU}^+$ cells were observed (Fig. 2g–i), providing evidence for CARN self-renewal. In particular, the percentage of BrdU^+ cells among $\text{Nkx3-1}^+ \text{YFP}^+$ cells, corresponding to CARNs in both the first and second regression, represents the percentage of CARNs undergoing a self-renewal division (24%, $n = 68$; Supplementary Table 2).

To assess long-term self-renewal, we examined the persistence of lineage-marked cells in *Nkx3-1*^{CreERT2/+}; *R26R-YFP/+* mice after four rounds of regression/regeneration (Fig. 2j). In these mice, YFP^+ cells represented 3.0% ($n = 21,559$) of the prostate epithelium, similar to the percentage observed after one round (Fig. 2k, l). The persistence of YFP^+ cells is consistent with the maintenance of a constant stem cell number during regeneration, as suggested by the

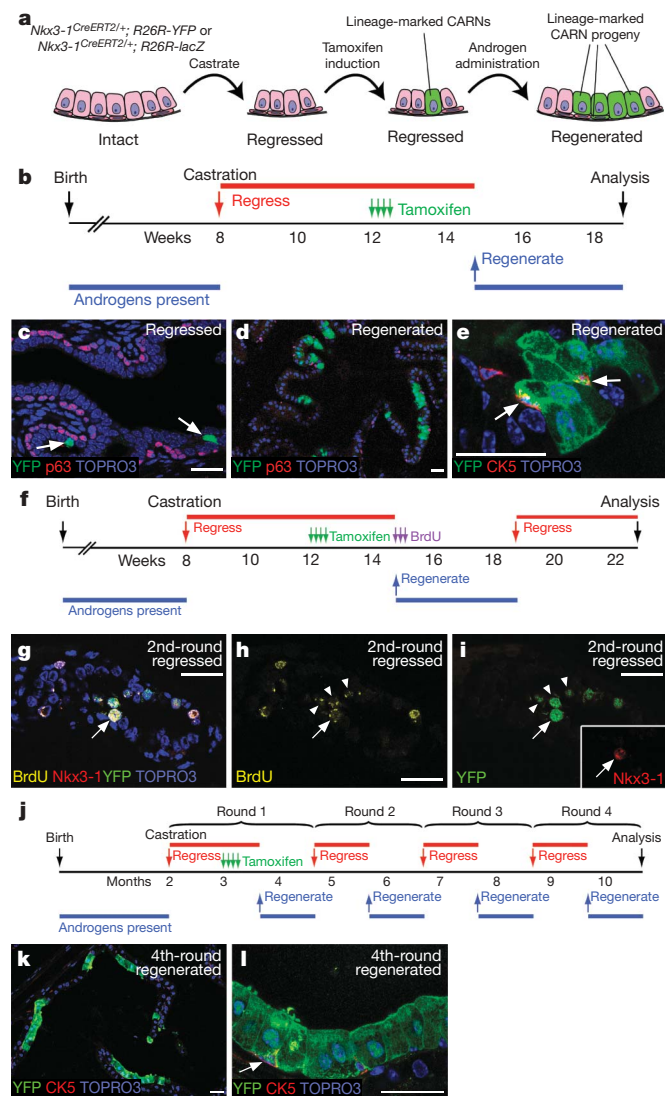


Figure 2 | Bipotentiality and self-renewal of CARNs in vivo. **a**, Strategy for lineage-marking experiment. **b**, Time-line for the experiment. **c**, YFP does not co-localize with p63 in lineage-marked cells of a castrated and tamoxifen-induced *Nkx3-1^{CreERT2/+}; R26R-YFP/+* anterior prostate. **d**, Clusters of YFP⁺ cells in a lineage-marked and regenerated prostate. **e**, Co-localization of YFP and CK5 in lineage-marked basal cells (arrows) of a regenerated prostate. **f**, Time-line for self-renewal experiment. **g–i**, Co-localization of Nkx3-1, YFP and BrdU immunostaining (arrow) in anterior prostate, shown as an overlay (**g**) and individual channels (**h**, **i**); YFP⁺ BrdU⁺ neighbours are indicated (arrowheads). **j**, Strategy for four-round serial regression/regeneration assay of long-term CARN self-renewal. **k**, **l**, Clusters of YFP⁺ cells in the lineage-marked prostate after four rounds of serial regression/regeneration. Scale bars, 25 μ m.

ability of the epithelium to undergo apparently unlimited serial regeneration^{5,6}, and supports the long-term self-renewal of lineage-marked CARNs.

Single-cell transplantation of CARNs

Next, we investigated whether CARNs could reconstitute prostate tissue in grafts generated from single or multiple lineage-marked CARNs (Fig. 3a and Supplementary Fig. 4). To examine single lineage-marked CARNs, we isolated individual YFP⁺ cells from suspensions of dissociated prostate cells, followed by recombination with rat urogenital mesenchyme cells and renal grafting in immunodeficient male mice (Supplementary Fig. 5a–f). The resulting grafts generated prostatic ducts with epithelial cells that were entirely YFP⁺ and that expressed luminal markers (E-cadherin, CK18 and androgen receptor), basal markers (p63 and CK5), or neuroendocrine

markers (synaptophysin) (Fig. 3b–h and Supplementary Fig. 5g, h); in particular, these ducts produced secretory proteins and expressed Nkx3-1, which is prostate-specific (Fig. 3c, i). Furthermore, we verified that the tissue formed in these grafts was unequivocally of mouse origin by nuclear morphology³¹ (Supplementary Fig. 6). Notably, the frequency of successful single-cell transplantation of lineage-marked YFP⁺ cells (37%, $n = 43$) was significantly greater than for the YFP[−] control (3%, $n = 31$; $P < 0.001$) (Fig. 3j).

Nkx3-1 regulates progenitor maintenance

Because *Nkx3-1* expression marks the CARN population, we next investigated whether *Nkx3-1* regulates progenitor maintenance and/or differentiation. First, we examined whether BrdU label-retaining cells (LRCs) might be affected by *Nkx3-1* inactivation, because in many tissues (but not all³²) such long-term growth-quiescent cells are enriched for progenitors^{33,34}. In the prostate, such LRCs can be identified by BrdU pulse-chase labelling during serial regression/regeneration⁶ (Fig. 4a). Under conditions in which 1.4% of epithelial cells ($n = 33,086$) retained BrdU-labelling at the fifth regression, 14.0% of CARNs ($n = 193$) were also BrdU-positive (Fig. 4b–d and Supplementary Table 3), indicating that a significant proportion of CARNs are also LRCs. Second, the percentage of LRCs in *Nkx3-1* mutants (0.3%, $n = 86,601$) was significantly less than in wild-type controls (0.8%, $n = 75,758$; $P = 0.003$) after five rounds of regression/regeneration (Fig. 4e–g and Supplementary Table 3), suggesting a decrease in prostate epithelial progenitors.

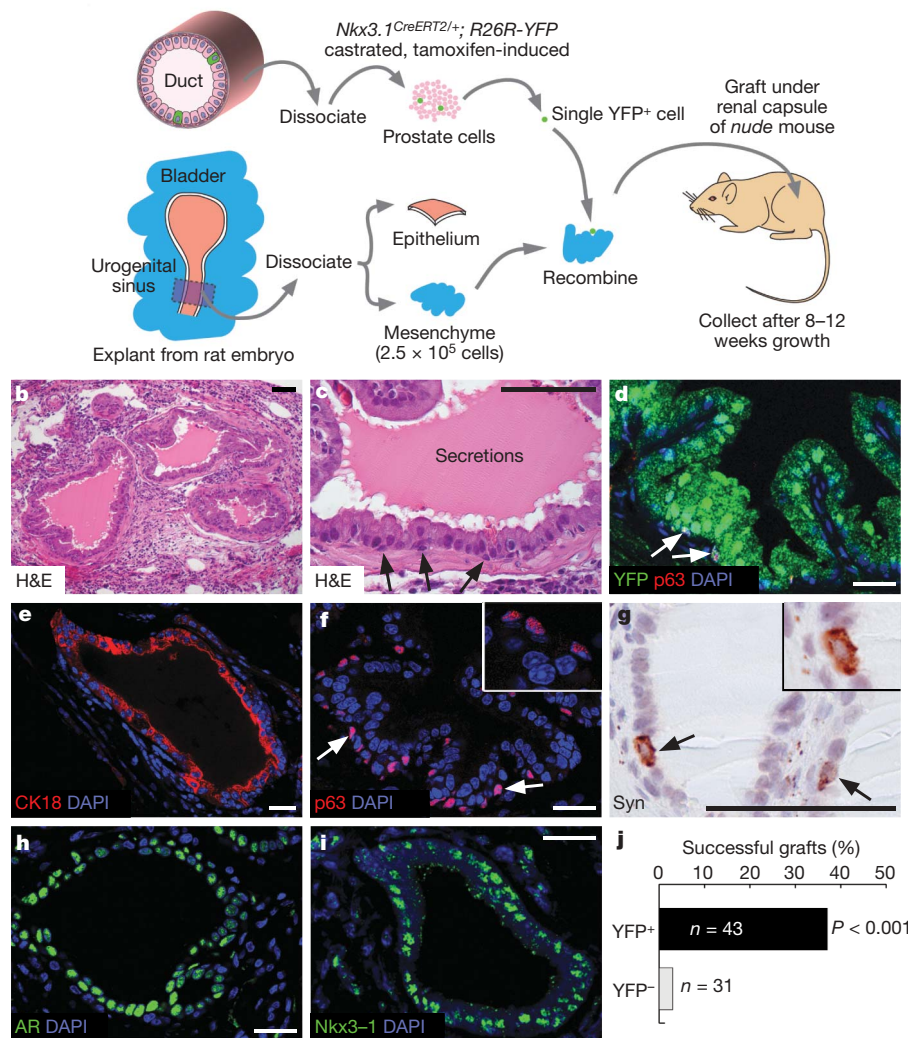
We also observed phenotypic alterations in *Nkx3-1* mutants after five rounds of serial regeneration, including reduced anterior prostate volume relative to wild-type controls (Fig. 4h). At the histological level, the characteristic hyperplasia and PIN phenotype of *Nkx3-1* homozygous ($n = 10$) as well as heterozygous ($n = 8$) mice was partially suppressed by five rounds of serial regeneration, whereas no abnormalities were observed in wild-type controls ($n = 9$) treated in parallel; similar results were observed after three rounds of serial regeneration (Supplementary Figs 7, 8 and Supplementary Table 4). Notably, the proliferative index of serially regenerated *Nkx3-1^{−/−}* mutants was similar to controls (Supplementary Fig. 8h, i), in contrast with the increased proliferation observed in intact *Nkx3-1* homozygotes²¹. Overall, these findings suggest that *Nkx3-1* is required for prostate stem cell maintenance during serial regression/regeneration.

CARNs are a cell of origin for cancer

We also investigated whether CARNs could represent a target of oncogenic transformation in prostate cancer, by examining the effects of CARN-specific deletion of the tumour suppressor gene *Pten*, an important regulator of the PI3-kinase/Akt signalling pathway that is frequently inactivated in human prostate cancer. For this purpose, we inducibly deleted *Pten* in the CARN population of castrated male mice carrying a conditional *Pten* allele³⁵ together with the inducible *Nkx3-1^{CreERT2}* allele (Fig. 5a). After androgen-mediated regeneration of the prostate, we observed rapid formation of high-grade PIN and carcinoma with evidence of microinvasion in the *Nkx3-1^{CreERT2/+}; Pten^{fllox/flox}* mice ($n = 6$), whereas control *Nkx3-1^{CreERT2/+}; Pten^{+/+}* mice were phenotypically normal ($n = 6$) (Fig. 5b–e). Notably, these PIN and carcinoma lesions showed increased proliferation and loss of basal cells, and displayed membrane-localized phosphorylated-Akt activity (Fig. 5f–m). These data indicate that the CARN population can act as a cell of origin for prostate cancer, and that the resulting carcinoma lesions have a luminal phenotype.

Discussion

Together with previous studies describing basal stem cells^{9,10,13}, our identification of CARNs as luminal stem cells indicates the existence of distinct non-overlapping stem cell populations in the prostate epithelium. Consequently, we can propose two general models for

a Explant from lineage-marked mouse prostate**Figure 3 | Generation of prostatic ducts in renal grafts by single lineage-marked CARNs.**

a, Strategy for tissue recombinant/renal graft analyses using a single YFP⁺ cell (or single YFP⁻ cell as a control). **b**, **c**, Haematoxylin and eosin (H&E) staining of prostatic ducts in a graft derived from a single YFP⁺ cell; note the presence of basal cells (arrows) and secretions (**c**). **d**, All epithelial cells in single-YFP⁺-derived duct express YFP, including p63⁺ basal cells (arrows). **e–g**, Expression of luminal marker CK18 (**e**), basal marker p63 (**f**), and neuroendocrine marker synaptophysin (Syn) (**g**) in ducts from single YFP⁺ cells. **h**, **i**, Expression of androgen receptor (AR) (**h**) and Nkx3-1 (**i**) confirm prostate identity of ducts. **j**, Summary of single-cell transplantation data. Scale bars, 25 μ m (**d–f**, **h**, **i**) and 50 μ m (**b**, **c**, **g**).

the lineage relationship between CARNs and a basal stem cell population. One possibility is that basal and luminal cell types may possess independent progenitors that have partially redundant stem cell activities (Fig. 6a). A second possibility is that CARNs represent facultative or 'potential' stem cells corresponding to transit-amplifying cells that acquire stem cell properties during regeneration/wound healing responses, as has been described in the mammalian testis

and pancreas^{36–38} (Fig. 6b). In this model, the facultative stem cells that drive prostate regeneration could be independent from the stem cells for prostate organogenesis, and might co-exist in the adult gland; such a dual progenitor system functions during *Drosophila* tracheal remodelling³⁹.

Furthermore, the observed defect in stem cell maintenance in *Nkx3-1* mutants suggests a functional role for *Nkx3-1* expression in CARNs.

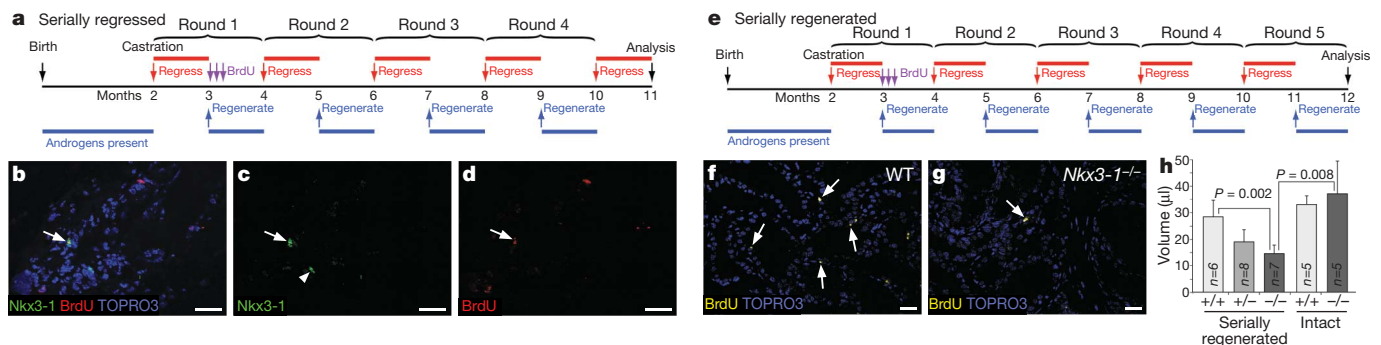


Figure 4 | *Nkx3-1* mutants display prostate epithelial defects in a serial regeneration assay. **a**, Time-line for analysis of LRCs. **b–d**, Overlap of CARNs with LRCs in a serially regressed prostate, shown as an overlay (**b**) and individual panels (**c**, **d**). Arrows in **b–d** indicate a *Nkx3-1⁺* BrdU⁺ cell; arrowhead in **c** indicates a CARN that is BrdU⁻. **e**, Time-line for serial regression/regeneration analyses. **f**, **g**, Decreased number of LRCs (arrows)

in *Nkx3-1^{-/-}* anterior prostate (**g**) relative to wild-type (WT) controls (**f**) after serial regeneration. **h**, Decreased volume of *Nkx3-1^{-/-}* anterior prostate relative to wild-type and *Nkx3-1^{+/-}* prostates after serial regeneration, and to intact wild-type and *Nkx3-1^{-/-}* prostates. Error bars correspond to one standard deviation. Scale bars, 25 μ m.

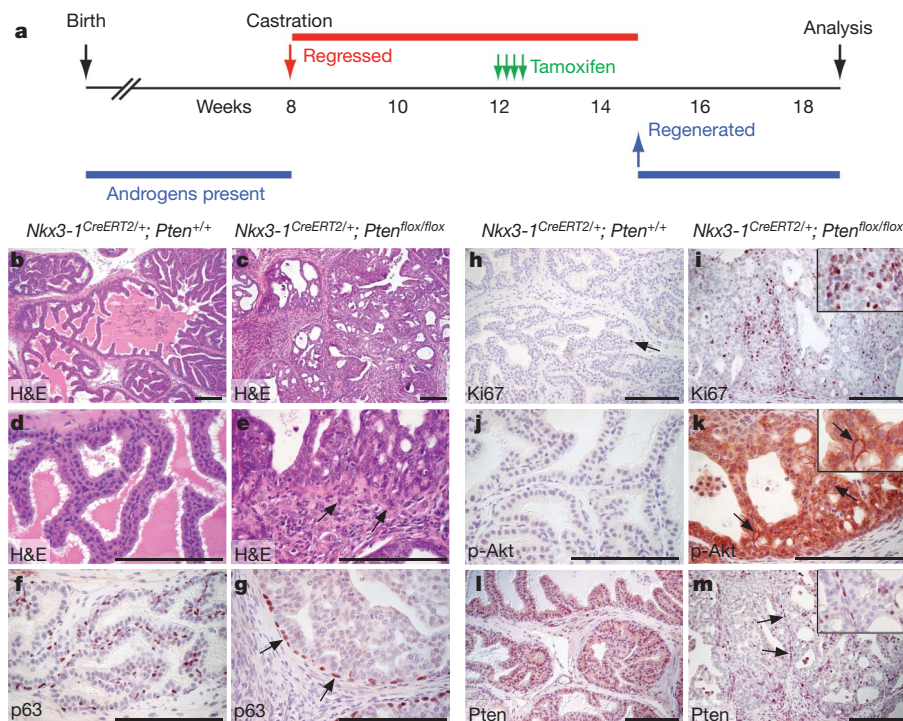


Figure 5 | The CARN population contains a cell type of origin for prostate cancer. **a**, Time-line for inducible conditional deletion of *Pten* in CARNs. **b–e**, H&E staining of anterior prostate from control *Nkx3-1*^{CreERT2/+}; *Pten*^{+/+} (**b**, **d**) and *Nkx3-1*^{CreERT2/+}; *Pten*^{flox/flox} (**c**, **e**) mice, shown at low-power (**b**, **c**) and high-power (**d**, **e**). The *Nkx3-1*^{CreERT2/+}; *Pten*^{flox/flox} prostate contains high-grade PIN/carcinoma lesions with local invasive epithelium (arrows, **e**). **f**, **g**, Detection of p63⁺ basal cells shows loss of basal

cells except at the periphery (arrows, **g**) of PIN/carcinoma lesions. **h**, **i**, Increased Ki67 immunostaining in PIN/carcinoma lesions. **j**, **k**, Phosphorylated-Akt (p-Akt) immunostaining with cell membrane localization (arrows, **k**) in PIN/carcinoma lesions. **l**, **m**, Pten immunostaining is ubiquitous in control *Nkx3-1*^{CreERT2/+}; *Pten*^{+/+} prostate epithelium, but is restricted to basal cells (arrows) and scattered luminal cells in induced *Nkx3-1*^{CreERT2/+}; *Pten*^{flox/flox} prostate. Scale bars, 100 μ m.

Thus, *Nkx3-1* inactivation might result in increased differentiation of CARNs and expansion of a proliferative transit-amplifying population (Supplementary Fig. 9). This interpretation is consistent with the finding that the duration of epithelial proliferation during prostate regeneration is prolonged in *Nkx3-1* mutants relative to wild type⁴⁰. The function of *Nkx3-1* in stem-cell maintenance may be direct, consistent with its feedback loop with the androgen receptor and its role in

prostate epithelial differentiation^{20,41}, or may be indirect, for example owing to increased oxidative damage with ageing⁴².

Finally, the importance of the stem cell compartment as a target of oncogenic transformation has been highlighted by studies showing that stem cell populations in the lung and colon are efficient cells of origin for cancer^{43,44}. In the case of prostate cancer, the identification of a castration-resistant stem cell population as a cell of origin also has implications for the onset of hormone-refractory disease. Thus, if oncogenic transformation of CARNs can result in the formation of a putative cancer stem cell, the eventual emergence of hormone-refractory disease may be prefigured through an initiating event during prostate carcinogenesis.

METHODS SUMMARY

The *Nkx3-1*^{CreERT2/+} allele was generated by gene targeting using standard techniques; the *Nkx3-1*-null mutant mice have been previously described²¹. *R26R-lacZ* and *Pten* conditional mutant mice were obtained from the Jackson Laboratory Induced Mutant Resource; the *R26R-YFP* mice were provided by F. Costantini. All lines were maintained on a hybrid C57BL/6-129/Sv strain background.

Castration of adult male mice was performed using standard techniques. For tamoxifen induction of Cre activity in mice containing *Nkx3-1*^{CreERT2/+}, mice were administered 9 mg per 40 g tamoxifen for 4 consecutive days. For prostate regeneration, physiological levels of testosterone (1.875 μ g h⁻¹) were administered for 4 weeks by subcutaneous implantation of mini-osmotic pumps (Alzet)⁴⁵. When included, BrdU (100 mg kg⁻¹) was administered once daily during the first 3 days of regeneration. For single-cell transplantation, single YFP⁺ cells were isolated by mouth-pipetting under epifluorescence illumination from a dissociated prostate cell suspension obtained from castrated and tamoxifen-induced *Nkx3-1*^{CreERT2/+}; *R26R-YFP/+* mice. A single YFP⁺ cell (or YFP⁻ cell as a control) was recombined with 2.5×10^5 rat urogenital sinus mesenchyme cells in a 10- μ l collagen pad, followed by transplantation under the kidney capsule of *nude* mice and collecting after 10–12 weeks.

Cryosections were stained with primary antibodies as listed in Supplementary Table 5, and counterstained with TOPRO3 or 4,6-diamidino-2-phenylindole

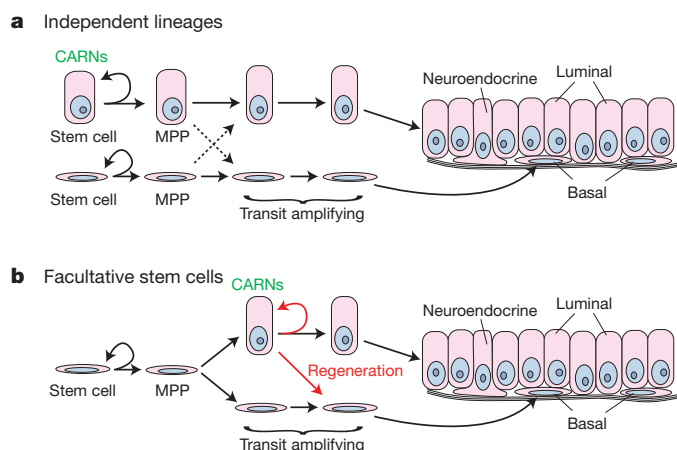


Figure 6 | Possible lineage relationships in the prostate epithelium. **a**, Independent stem cells for basal and luminal epithelium may give rise to differentiated cell types through multipotent progenitors (MPP) and transit-amplifying progenitors, with some bipotentiality (dashed arrows). In this model, CARNs would correspond to the luminal stem cells. **b**, Alternatively, stem cells for prostate organogenesis may be basal, but luminal transit-amplifying cells, including CARNs, can acquire stem cell properties during regeneration (red arrows), thus acting as facultative or potential stem cells.

(DAPI) (Invitrogen/Molecular Probes). Secondary antibodies were labelled with Alexa Fluor 488, 555 or 594 (Invitrogen/Molecular Probes). Immunofluorescence staining was imaged using a Leica TCS5 spectral confocal microscope. Cell counting was performed manually using confocal photomicrographs with at least three animals for each experiment or genotype analysed.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 April 2008; accepted 5 August 2009.

Published online 9 September 2009.

- Abate-Shen, C. & Shen, M. M. Molecular genetics of prostate cancer. *Genes Dev.* **14**, 2410–2434 (2000).
- English, H. F., Santen, R. J. & Isaacs, J. T. Response of glandular versus basal rat ventral prostatic epithelial cells to androgen withdrawal and replacement. *Prostate* **11**, 229–242 (1987).
- Evans, G. S. & Chandler, J. A. Cell proliferation studies in the rat prostate: II. The effects of castration and androgen-induced regeneration upon basal and secretory cell proliferation. *Prostate* **11**, 339–351 (1987).
- Sugimura, Y., Cunha, G. R. & Donjacour, A. A. Morphological and histological study of castration-induced degeneration and androgen-induced regeneration in the mouse prostate. *Biol. Reprod.* **34**, 973–983 (1986).
- Isaacs, J. T. in *Benign Prostatic Hyperplasia* (eds Rodgers C. H. et al.) 85–94 (Department of Health and Human Services, 1985).
- Tsujimura, A. et al. Proximal location of mouse prostate epithelial stem cells: a model of prostatic homeostasis. *J. Cell Biol.* **157**, 1257–1265 (2002).
- Lawson, D. A. & Witte, O. N. Stem cells in prostate cancer initiation and progression. *J. Clin. Invest.* **117**, 2044–2050 (2007).
- Senoo, M., Pinto, F., Crum, C. P. & McKeon, F. p63 is essential for the proliferative potential of stem cells in stratified epithelia. *Cell* **129**, 523–536 (2007).
- Lawson, D. A., Xin, L., Lukacs, R. U., Cheng, D. & Witte, O. N. Isolation and functional characterization of murine prostate stem cells. *Proc. Natl Acad. Sci. USA* **104**, 181–186 (2007).
- Richardson, G. D. et al. CD133, a novel marker for human prostatic epithelial stem cells. *J. Cell Sci.* **117**, 3539–3545 (2004).
- Burger, P. E. et al. Sca-1 expression identifies stem cells in the proximal region of prostatic ducts with high capacity to reconstitute prostatic tissue. *Proc. Natl Acad. Sci. USA* **102**, 7180–7185 (2005).
- Xin, L., Lawson, D. A. & Witte, O. N. The Sca-1 cell surface marker enriches for a prostate-regenerating cell subpopulation that can initiate prostate tumorigenesis. *Proc. Natl Acad. Sci. USA* **102**, 6942–6947 (2005).
- Goldstein, A. S. et al. Trop2 identifies a subpopulation of murine and human prostate basal cells with stem cell characteristics. *Proc. Natl Acad. Sci. USA* **105**, 20882–20887 (2008).
- Leong, K. G., Wang, B. E., Johnson, L. & Gao, W. Q. Generation of a prostate from a single adult stem cell. *Nature* **456**, 804–808 (2008).
- Kurita, T., Medina, R. T., Mills, A. A. & Cunha, G. R. Role of p63 and basal cells in the prostate. *Development* **131**, 4955–4964 (2004).
- Kasper, S. Stem cells: the root of prostate cancer? *J. Cell. Physiol.* **216**, 332–336 (2008).
- Wang, S. et al. Pten deletion leads to the expansion of a prostatic stem/progenitor cell subpopulation and tumor initiation. *Proc. Natl Acad. Sci. USA* **103**, 1480–1485 (2006).
- Brisanzio, C. & Signoretti, S. p63 in prostate biology and pathology. *J. Cell. Biochem.* **103**, 1354–1368 (2008).
- Humphrey, P. A. Diagnosis of adenocarcinoma in prostate needle biopsy tissue. *J. Clin. Pathol.* **60**, 35–42 (2007).
- Abate-Shen, C., Shen, M. M. & Gelmann, E. Integrating differentiation and cancer: the *Nkx3.1* homeobox gene in prostate organogenesis and carcinogenesis. *Differentiation* **76**, 717–727 (2008).
- Bhatia-Gaur, R. et al. Roles for *Nkx3.1* in prostate development and cancer. *Genes Dev.* **13**, 966–977 (1999).
- Abdulkadir, S. A. et al. Conditional loss of *Nkx3.1* in adult mice induces prostatic intraepithelial neoplasia. *Mol. Cell. Biol.* **22**, 1495–1503 (2002).
- Kim, M. J. et al. *Nkx3.1* mutant mice recapitulate early stages of prostate carcinogenesis. *Cancer Res.* **62**, 2999–3004 (2002).
- Chen, H., Mutton, L. N., Prins, G. S. & Bieberich, C. J. Distinct regulatory elements mediate the dynamic expression pattern of *Nkx3.1*. *Dev. Dyn.* **234**, 961–973 (2005).
- Scivolino, P. J. et al. Tissue-specific expression of murine *Nkx3.1* in the male urogenital system. *Dev. Dyn.* **209**, 127–138 (1997).
- Bieberich, C. J., Fujita, K., He, W.-W. & Jay, G. Prostate-specific and androgen-dependent expression of a novel homeobox gene. *J. Biol. Chem.* **271**, 31779–31782 (1996).
- Feil, R., Wagner, J., Metzger, D. & Chambon, P. Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem. Biophys. Res. Commun.* **237**, 752–757 (1997).
- Indra, A. K. et al. Temporally-controlled site-specific mutagenesis in the basal layer of the epidermis: comparison of the recombinase activity of the tamoxifen-inducible Cre-ER(T) and Cre-ER(T2) recombinases. *Nucleic Acids Res.* **27**, 4324–4327 (1999).
- Srinivas, S. et al. Cre reporter strains produced by targeted insertion of *EYFP* and *ECFP* into the *ROSA26* locus. *BMC Dev. Biol.* **1**, 4 (2001).
- Soriano, P. Generalized *lacZ* expression with the *ROSA26* Cre reporter strain. *Nature Genet.* **21**, 70–71 (1999).
- Cunha, G. R. & Vanderslice, K. D. Identification in histological sections of species origin of cells from mouse, rat and human. *Stain Technol.* **59**, 7–12 (1984).
- Kiel, M. J. et al. Haematopoietic stem cells do not asymmetrically segregate chromosomes or retain BrdU. *Nature* **449**, 238–242 (2007).
- Bickenbach, J. R. & Holbrook, K. A. Label-retaining cells in human embryonic and fetal epidermis. *J. Invest. Dermatol.* **88**, 42–46 (1987).
- Cotsarelis, G., Cheng, S. Z., Dong, G., Sun, T. T. & Lavker, R. M. Existence of slow-cycling limbal epithelial basal cells that can be preferentially stimulated to proliferate: implications on epithelial stem cells. *Cell* **57**, 201–209 (1989).
- Groszer, M. et al. Negative regulation of neural stem/progenitor cell proliferation by the Pten tumor suppressor gene *in vivo*. *Science* **294**, 2186–2189 (2001).
- Nakagawa, T., Nabeshima, Y. & Yoshida, S. Functional identification of the actual and potential stem cell compartments in mouse spermatogenesis. *Dev. Cell* **12**, 195–206 (2007).
- Xu, X. et al. Beta cells can be generated from endogenous progenitors in injured adult mouse pancreas. *Cell* **132**, 197–207 (2008).
- Barroca, V. et al. Mouse differentiating spermatogonia can generate germinal stem cells *in vivo*. *Nature Cell Biol.* **11**, 190–196 (2009).
- Weaver, M. & Krasnow, M. A. Dual origin of tissue-specific progenitor cells in *Drosophila* tracheal remodeling. *Science* **321**, 1496–1499 (2008).
- Magee, J. A., Abdulkadir, S. A. & Milbrandt, J. Haploinsufficiency at the *Nkx3.1* locus: a paradigm for stochastic, dosage-sensitive gene regulation during tumor initiation. *Cancer Cell* **3**, 273–283 (2003).
- Lei, Q. et al. *NKX3.1* stabilizes p53, inhibits AKT activation, and blocks prostate cancer initiation caused by PTEN loss. *Cancer Cell* **9**, 367–378 (2006).
- Ouyang, X., DeWeese, T. L., Nelson, W. G. & Abate-Shen, C. Loss-of-function of *Nkx3.1* promotes increased oxidative damage in prostate carcinogenesis. *Cancer Res.* **65**, 6773–6779 (2005).
- Kim, C. F. et al. Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell* **121**, 823–835 (2005).
- Barker, N. et al. Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457**, 608–611 (2009).
- Banach-Petrosky, W. et al. Prolonged exposure to reduced levels of androgen accelerates prostate cancer progression in *Nkx3.1*; *Pten* mutant mice. *Cancer Res.* **67**, 9089–9096 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Kim for her initial observations on *Nkx3.1* expression in the regressed prostate, and C. Cordon-Cardo, E. Gelmann, C. Mendelsohn and B. Reizis for comments on the manuscript. We are also grateful to C. Bieberich, M. Capocchi, P. Chambon and F. Costantini for providing mice and reagents. This work was supported by grants from the NIH (C.A.-S. and M.M.S.), DOD Prostate Cancer Research Program (K.D.E., C.A.-S. and M.M.S.), and the NCI Mouse Models of Human Cancer Consortium.

Author Contributions X.W., M.K.-D., K.D.E., C.A.-S. and M.M.S. designed experiments, Y.P.-H. and S.M.P. generated mouse reagents, X.W., M.K.-D., K.D.E., D.W., H.Y. and M.V.H. performed experiments, and X.W., M.K.-D., C.A.-S. and M.M.S. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.M.S. (mshen@columbia.edu).

METHODS

Gene targeting and genotyping. The *Nkx3-1^{CreERT2/+}* allele was generated by gene targeting using standard techniques⁴⁶. The targeting vector was generated using a 5' arm corresponding to a 3.5-kilobase (kb) PCR fragment from a *Nkx3-1* genomic clone²¹ up to the translation initiation site of *Nkx3-1*, and a 3' arm corresponding to a 4.0-kb PCR fragment of genomic sequence (Supplementary Fig. 2a). The positive-selection cassette corresponded to the self-excising ACE-Cre/PolII-neo selection cassette from the pACN vector⁴⁷, whereas negative selection was provided by the PGK-tk cassette from the pPNT vector⁴⁸. Primers for generating the 5' arm were 5'-ACCGGAATTCTCCGCTGCGCGCCGCTTTT GC-3' and 5'-ACCCAAGCTTCATGCCTGCAGGTCGGAGGCC-3'. Primers to amplify the 3' arm were 5'-CTAGTCTAGAGCGGCTCACCTCCTTC CTCA-3' and 5'-CTAGTCTAGAGGATGGCAGGAGAGGTCAGTGC-3'. The gap between the 5' and 3' arms is approximately 80 base pairs (bp), such that the 3' arm contains most of exon 1 together with intron 1, exon 2 and 200 bp of genomic sequence 3' of the transcription termination site. The pGS-CreER^{T2} vector²⁸ was provided by P. Chambon, and was modified by the insertion of a 65-bp intron from ACE-Cre⁴⁷ between the PshAI and ClaI sites of the CreER^{T2} sequence. Culture and transfection of mouse embryonic stem (ES) cells followed standard protocols⁴⁶. Homologous recombinants in TC1 ES cells⁴⁹ were selected by positive-negative selection followed by Southern blot screening. One out of two-hundred-and-sixty clones analysed was properly targeted, and this clone was used to generate germline chimaeras.

Mouse genotyping. Genotyping for the *Nkx3-1^{CreERT2}* allele was performed by Southern blotting or by PCR using tail genomic DNA. Primers for PCR genotyping were as follows: for the *Nkx3-1* wild-type allele, 5'-CTCCGCTACCCTA AGCATCC-3' and 5'-GACACTGTCACTATTACTTGGACC-3', which amplifies a region deleted in the targeting vector; and for the *Nkx3-1^{CreERT2}* allele, 5'-CAGATGGCGCGCAACACC-3' and 5'-GCGCGGTCTGCGAGTAAAC-3'.

The primers for genotyping *Nkx3-1* mutant mice were 5'-GCCAACCTGCCT CAATCACTAAGG-3' (wild-type *Nkx3-1* forward), 5'-TTCCACATACACTTC ATTCTCAGT-3' (mutated forward), and 5'-GCCAACCTGCCTCAATCACTA AGG-3' (wild-type and mutated reverse). The primers for genotyping the *R26R-lacZ* Cre-reporter were 5'-CCGCGCTGTACTGGAGGCTGAAG-3' (forward) and 5'-ATACTGCACCGGGCGGAAGGAT-3' (reverse). Primers for genotyping the *Pten* conditional (*Pten^{fllox}*) allele were 5'-ACTCAAGCAGGGATGAGC-3' (forward) and 5'-GTCATCTTCACTTAGCCATTGG-3' (reverse). Primers for genotyping the *R26R-YFP* mice were 5'-GCGAAGAGTTTGTCTCAACC-3' (mutated forward), 5'-GGAGCGGGAGAAATGGATATG-3' (wild-type forward) and 5'-AAAGTCGCTCTGAGTTGTAT-3' (wild-type and mutated reverse).

Mouse procedures. Castration of adult male mice was performed using standard techniques⁵⁰. After castration at 8 weeks of age, mice were allowed to regress for 4 weeks to reach the fully involuted state. For tamoxifen induction of Cre activity in mice containing the *Nkx3-1^{CreERT2}* allele, mice were administered 9 mg per 40 g tamoxifen (Sigma) suspended in corn oil, or vehicle alone for negative controls, by intraperitoneal injection or oral gavage once daily for 4 consecutive days, followed by a chase period of 14 days.

For prostate regeneration, testosterone (Sigma) was dissolved at 25 mg ml⁻¹ in 100% ethanol and diluted in PEG-400 to a final concentration of 7.5 mg ml⁻¹. Testosterone was administered for 4 weeks at a rate of 1.875 µg h⁻¹ delivered by subcutaneous implantation of mini-osmotic pumps (Alzet); this regimen yields physiological levels of serum testosterone⁴⁵. When included, BrdU (100 mg kg⁻¹) (Sigma) was also administered by intraperitoneal injection once daily during the first 3 days of regeneration to label proliferating cells. After regeneration of the prostate, mice could be euthanized for analysis, or deprived of androgens by pump removal, returning to the regressed state after 4 more weeks. At this point, mice were either euthanized for analysis, or osmotic pumps could be reimplanted for further rounds of serial regression/regeneration.

For tissue recombination and renal grafting, prostate tissues (corresponding to the combined anterior, dorsolateral and ventral lobes) were dissected and minced to small clumps, followed by enzymatic dissociation with 0.2% collagenase I (Invitrogen) in DMEM media with 10% FBS for 90 min. Dissociated tissue was passed sequentially through 21-, 23- and 26-gauge needles followed by a 40-µm cell strainer to obtain single-cell suspensions. The resulting cells were assessed for viability by trypan blue exclusion and counted. For grafts containing large numbers of epithelial cells, as in Supplementary Fig. 4, 2.5 × 10⁵ dissociated prostate cells obtained from castrated and tamoxifen-induced *Nkx3-1^{CreERT2/+}*; *R26R-YFP/+* mice were mixed with 2.5 × 10⁵ dissociated urogenital sinus mesenchyme (UGM) cells from embryonic day (E) 18.0 rat embryos. UGM cells were obtained from dissected urogenital sinus that was treated for 30 min in 1% trypsin, followed by mechanical dissociation and treatment with 0.1% collagenase B (Roche) for 30 min at 37 °C, and washing in PBS. Pelleted cell mixtures were resuspended in 10 µl of 1:5 collagen:setting buffer (10× Earle's Balanced Salt

Solution (Life Technologies), 0.2 M NaHCO₃ and 50 mM NaOH), and gelatinized in 37 °C for 20 min. Tissue recombinants were cultured in DMEM media with 10% FBS supplemented with 10⁻⁷ M dihydrotestosterone (DHT) overnight, followed by transplantation under the kidney capsules of nude mice. Grafts were collected after 4–8 weeks of growth for analysis.

For single-cell grafts, as in Fig. 3, a single YFP⁺ (or YFP⁻ cell as a control) was isolated from the dissociated cell suspension from castrated and tamoxifen-induced *Nkx3-1^{CreERT2/+}*; *R26R-YFP/+* prostates by mouth-pipetting under epifluorescence illumination on an Olympus IX51 inverted microscope with DP71 camera. This single cell was then recombined with 2.5 × 10⁵ dissociated UGM cells obtained from E18.0 rat embryos, and cultured and grafted as above. Grafts were collected after 10–12 weeks of growth for analysis, and imaged under epifluorescence on an Olympus SZX16 stereomicroscope with DP71 camera. The resulting graft tissue was analysed for YFP and other marker expression as described later, and counterstained with DAPI for visualization of nuclear morphology at high-power to distinguish mouse from rat nuclei³¹.

Grafts recovered from transplantation of a single lineage-marked YFP⁺ cell (*n* = 16 out of 43, 37%) were confirmed to be of mouse origin by YFP expression and nuclear morphology (Supplementary Fig. 6), whereas the single graft (*n* = 1 out of 31, 3%) arising from a YFP⁻ cell was confirmed to be of mouse origin by nuclear morphology. Generation of prostatic ducts by a YFP⁻ cell might result from a CARN that was not lineage-marked by tamoxifen-induction, which is inefficient, or alternatively from a distinct stem cell type in the prostate epithelium. We note that a considerable percentage (*n* = 13 out of 74, 18%) of the grafts contained ducts of rat origin, which are not included in Fig. 3j. These rat ducts probably arise from rat urogenital epithelial cells that are difficult to completely dissociate from the urogenital mesenchyme used in the graft, and can populate the graft under conditions in which the epithelial contribution is limiting.

For histological and immunofluorescence analysis, individual prostate lobes or renal grafts were dissected, and then fixed in 4% paraformaldehyde for subsequent cryoembedding in OCT compound (Sakura), or fixed in 10% formalin followed by paraffin embedding. The volume of dissected anterior prostate lobes was determined by physical displacement of known volumes of PBS solution in 0.5-ml centrifuge tubes.

Histology and immunostaining. H&E staining was performed using standard protocols on 6-µm paraffin sections. β-galactosidase staining was performed using 12-µm cryosections, which were incubated in staining solution (0.1 M PBS, 1.3 mM MgCl₂, 1 mg ml⁻¹ X-gal, 0.02% Nonidet P-40, 5 mM K₄Fe(CN)₆, 5 mM K₃Fe(CN)₆ and 0.01% Na-deoxycholate) for 3 h or overnight, followed by fixation in 10% formalin for 2–5 h. Direct visualization of YFP was performed after washing 10-µm cryosections in PBST (PBS with 0.1% Triton X-100) three times, incubation with TOPRO3 (1:1,000 diluted in PBST) (Invitrogen/Molecular Probes) for 30 min, and mounting with VECTASHIELD mounting medium (Vector Labs), which contains DAPI.

For immunohistochemical staining, 6-µm paraffin sections were deparaffinized in xylene, followed by antigen retrieval through boiling in antigen unmasking solution (Vector Labs). Slides were blocked in 10% normal serum or with blocking reagents provided in the M.O.M. ('Mouse-on-Mouse') immunodetection kit (Vector Labs) for mouse primary antibodies, then incubated with primary antibodies overnight at 4 °C or room temperature. Primary antibodies and dilutions used are listed in Supplementary Table 5. Secondary antibodies were obtained from Vectastain ABC kits (Vector Labs) and diluted 1:250 or 1:500. The signal was enhanced using the Vectastain ABC system and visualized with the NovaRed Substrate Kit (Vector Labs). The slides were counterstained with Harris Modified Hematoxylin (diluted 1:4 in H₂O) (Fisher Scientific) and mounted with Clearmount (American Master*Tech Scientific). Immunohistochemical staining was imaged using a Nikon Eclipse E800 microscope equipped with a Nikon DXM1200 digital camera.

Immunofluorescence staining was performed on either 6-µm paraffin sections or 10-µm cryosections, which were incubated in 3% H₂O₂ and Antigen Unmasking Solution (Vector Labs). Primary antibodies and dilutions used are listed in Supplementary Table 5. Slides were incubated with 10% normal goat serum (Vector Labs) or donkey serum (Sigma) and with primary antibodies diluted in the 10% normal goat or donkey serum overnight at 4 °C or room temperature. Slides then were incubated with secondary antibodies (diluted 1:500 in PBST) labelled with Alexa Fluor 488, 555, or 594 (Invitrogen/Molecular Probes). Detection of *Nkx3-1*, green fluorescent protein (GFP) and Cre was enhanced using tyramide amplification (Invitrogen/Molecular Probes) by incubation of slides with horseradish peroxidase (HRP)-conjugated secondary antibody (1:100 dilution) (Invitrogen/Molecular Probes), followed by incubation with tyramide 488 or 555 for 6 min. Sections were counterstained with TOPRO3 or TOTO3 (diluted 1:1,000 in PBST) (Invitrogen/Molecular Probes) to visualize nuclei, and mounted with VECTASHIELD mounting medium (Vector Labs),

which contains DAPI. Immunofluorescence staining was imaged using a Leica TCS5 spectral confocal microscope.

Quantification and statistics. To calculate the number of CARNs in the regressed mouse prostate, we determined that there are an average of 112,000 total cells ($n = 5$ animals; all lobes combined), of which 59% are epithelial as determined by immunoreactivity for the pan-epithelial marker CD24 (ref. 9). Because 0.7% of epithelial cells in the regressed prostate are CARNs, there are approximately 460 CARNs in the total prostate. To determine the number of lineage-marked cells in the regressed prostate, we visualized 320 live YFP⁺ cells in dissociated prostate tissue (all lobes combined) from five castrated lineage-marked *Nkx3-1^{CreERT2/+}; R26R-YFP/+* mice, for a total of 64 YFP⁺ live cells/mouse. For the experiment in Supplementary Fig. 4, we performed two re-combinations from these dissociated prostate cells, so that there were approximately 160 live YFP⁺ cells used in each graft.

For immunostaining experiments, cell numbers were counted manually using confocal $\times 40$ and $\times 63$ photomicrographs. Statistical analyses were performed

using a two-sample t -test, χ^2 test, or Fisher's Exact test as appropriate. At least three animals for each experiment or genotype were analysed.

46. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. *Manipulating the Mouse Embryo: A Laboratory Manual* Chs 8–11 359–506 (Cold Spring Harbor Laboratory Press, 2003).
47. Bunting, M., Bernstein, K. E., Greer, J. M., Capecchi, M. R. & Thomas, K. R. Targeting genes for self-excision in the germ line. *Genes Dev.* **13**, 1524–1528 (1999).
48. Tybulewicz, V. L., Crawford, C. E., Jackson, P. K., Bronson, R. T. & Mulligan, R. C. Neonatal lethality and lymphopenia in mice with a homozygous disruption of the *c-abl* proto-oncogene. *Cell* **65**, 1153–1163 (1991).
49. Deng, C., Wynshaw-Boris, A., Zhou, F., Kuo, A. & Leder, P. Fibroblast growth factor receptor 3 is a negative regulator of bone growth. *Cell* **84**, 911–921 (1996).
50. Gao, H., Ouyang, X., Banach-Petrosky, W. A., Shen, M. M. & Abate-Shen, C. Emergence of androgen independence at early stages of prostate cancer progression in *Nkx3.1; Pten* mice. *Cancer Res.* **66**, 7929–7933 (2006).

Seismic evidence for the loss of stellar angular momentum before the white-dwarf stage

S. Charpinet¹, G. Fontaine² & P. Brassard²

White-dwarf stars represent the final products of the evolution of some 95% of all stars¹. If stars were to keep their angular momentum throughout their evolution, their white-dwarf descendants, owing to their compact nature, should all rotate relatively rapidly, with typical periods of the order of a few seconds. Observations of their photospheres show, in contrast, that they rotate much more slowly, with periods ranging from hours to tens of years^{2–5}. It is not known, however, whether a white dwarf could ‘hide’ some of its original angular momentum below the superficial layers, perhaps spinning much more rapidly inside than at its surface. Here we report a determination of the internal rotation profile of a white dwarf using a method based on asteroseismology. We show that the pulsating white dwarf PG 1159–035 rotates as a solid body (encompassing more than 97.5% of its mass) with the relatively long period of 33.61 ± 0.59 h. This implies that it has lost essentially all of its angular momentum, thus favouring theories which suggest important angular momentum transfer and loss in evolutionary phases before the white-dwarf stage^{6,7}.

PG 1159–035 is a relatively bright ($V = 14.87$) white dwarf showing multi-periodic luminosity variations caused by gravity-wave pulsational instabilities. It belongs to one of the four known categories of pulsating white dwarfs, the GW Vir stars, which are very hot objects ($T_{\text{eff}} \approx 80,000\text{--}170,000$ K) with atmospheres made of a mixture of helium, carbon and oxygen in roughly comparable proportions⁸. It is one of the best-studied pulsating white dwarfs, having been the subject of several multi-site observational campaigns⁹. These campaigns have revealed a rich pulsation spectrum made of dipole and quadrupole modes with periods ranging from 350.8 s to 988.1 s. A detailed seismic analysis of these pulsations¹⁰, based on purely spherically symmetric (non-rotating) stellar models, has led to a successful deciphering of the information contained in the periods detected in PG 1159–035. From this, the global structural parameters of PG 1159–035 have been inferred, including its chemical stratification as a function of depth, a property that can only be deduced from asteroseismology. In addition, the pulsations detected in PG 1159–035 have each been formally identified in terms of the indices k (radial order) and l (degree) that define a normal mode of oscillation in a non-rotating star.

The available data on PG 1159–035 also reveal that 17 of the principal pulsation modes exhibit a fine structure such that they are made of two to five components very closely spaced in period (see Supplementary Information for details). For example, the largest amplitude mode in PG 1159–035 shows a triplet structure with components having periods of 516.061 ± 0.001 s, 517.175 ± 0.002 s and 518.292 ± 0.002 s. Such a multiplet structure is best interpreted in terms of the phenomenon of rotational splitting in asteroseismology¹¹. The latter is related to the fact that a pulsation mode in a non-rotating star is $(2l + 1)$ -fold degenerate in period owing to the spherical symmetry of the system. Rotation lifts this degeneracy because it destroys

the spherical symmetry, and a formally degenerate mode may then show its $2l + 1$ components. Hence, there is information about the rotation state of a pulsating star encoded in the period spacings between the components of multiplet structures. We exploit this information here in the case of PG 1159–035.

The first step is to assume that (1) PG 1159–035 is a slow rotator, that is, its rotation period is much larger than the observed pulsation periods (a hypothesis that can easily be verified a posteriori), and (2) it rotates as a rigid body (an assumption that is tested below with the method that we developed¹² and apply here). Under these circumstances, pulsation theory allows the calculations of period splittings for each of the identified modes through a perturbation method¹¹. The unperturbed reference state is a model similar to the seismic spherical model of PG 1159–035 referred to above¹⁰. The calculated period spacings may then be compared with the observed multiplet structures, and the assumed rotation period is varied until a best global match between the calculated and observed spacings is obtained. To do this, we used a merit function, S^2 , inspired from a standard χ^2 approach, representing the weighted sum of the squared differences between the 31 observed period spacings and their expected counterparts, given a rotation period. This goodness-of-fit function was minimized by varying the input rotation period. The result of this optimization procedure is shown in Fig. 1, which illustrates a very well defined minimum in the merit function, corresponding to a rotation period of 33.61 ± 0.59 h for PG 1159–035. Figure 2 displays a comparison of the observed period spacings with those of the best-fitting rotating model. We note that the inferred rotation period is much larger than any of the pulsation periods observed in PG 1159–035, and this verifies, after the fact, our first initial assumption. We next tested the assumption of solid-body rotation.

We constructed test stellar models made of an outer zone assumed to rotate rigidly with the period inferred just above, and an inner zone, also assumed to rotate as a solid body, but with an arbitrary period that we vary. The pulsations in a star are global modes (they may be thought of as standing waves in the radial direction, with boundary conditions to be respected at the surface and at the centre), so the values of the calculated period spacings due to rotation depend on the whole vertical structure. Hence, by varying the depth location of the boundary between the upper zone and the lower zone, the values of the theoretical period spacings change. Our problem is then one of optimization in 2D space: we seek to minimize the merit quantity S^2 as a function of both the rotation period of the inner zone and the depth of the transition layer between the upper and lower zones. If PG 1159–035 really rotates as a solid body, then the S^2 function in a depth versus inner-zone period diagram should show contours taking the form of a vertical valley centred on the period of 33.61 h. This is precisely what Fig. 3 illustrates. The solid white curve

¹Laboratoire d'Astrophysique de Toulouse-Tarbes, Université de Toulouse, CNRS, 14 avenue E. Belin, 31400 Toulouse, France. ²Département de Physique, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, Québec H3C 3J7, Canada.

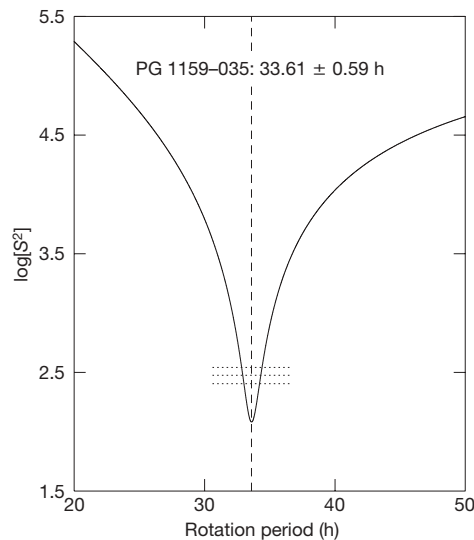


Figure 1 | Result of the optimization procedure under the hypothesis that PG 1159–035 rotates as a rigid body. This shows the behaviour of the merit function S^2 in terms of the assumed rotation period. The merit function exhibits a very well defined minimum, corresponding to a rotation period of 33.61 h as indicated by the dashed vertical line. The dotted horizontal lines correspond, from bottom to top, to the 1, 2 and 3σ limits. The stellar model that we used in these calculations was inspired from the seismic model inferred from an independent analysis of the pulsation properties of PG 1159–035 (ref. 10). The latter indicates that our target star has an effective temperature of about 128,000 K, a total mass of 0.565 times that of the Sun, a radius of 0.024 times the solar radius, and a total luminosity some 141 times that of the Sun.

in the figure indicates the minimum of the valley and, hence, the best solution for the rotation period of the inner zone at a given depth. The dotted curves give the 1, 2, and 3σ contours about the optimal solution. The fact that these contours flatten out at very large depths indicates that the available gravity waves progressively lose their usefulness as seismic probes at these depths. Notwithstanding this limitation, Fig. 3 confirms the hypothesis that PG 1159–035 rotates wholly as a solid body because only 2.5% of the total mass of the star is contained within the inner 10% of the radial coordinate where the method loses its sensitivity. This is, to our knowledge, the first time that the internal rotation profile of a white dwarf has been measured. We note that at least one published theoretical study¹³ suggests explicitly that isolated white dwarfs should rotate rigidly and not differentially, in agreement with our finding here, although the argument was made for cooler stars.

We emphasize that our method rests on the global nature of gravity-mode pulsations and can be used to verify whether a star rotates as a solid body or not. It is to be distinguished from inversion techniques used in helioseismology which allow the precise determination of the local rotation rate as a function of depth and latitude in the Sun^{14,15}. For these techniques to work, however, thousands to tens of thousands of periods must be known¹⁶, a luxury that asteroseismology in general simply cannot afford. We experimented with various methods for sounding the inner rotation profile of PG 1159–035 beyond the simple two-zone approach just described, and we found the same results. For instance, we assumed a linear law (in the radial coordinate) for the rotation profile by specifying the rotation period at the surface, P_s , and the rotation period at the centre, P_c . We repeated our two-dimensional optimization procedure, this time in P_s – P_c space, and we found an optimal solution corresponding to $P_s = P_c = 33.61$ h within the uncertainties, thus corroborating the case of rigid rotation in PG 1159–035. Our method can readily be applied to many other pulsating white dwarfs with available rotation data (period splittings), and we have begun this work. We note that GW Vir stars offer the best potential for internal sounding of this kind because gravity modes keep

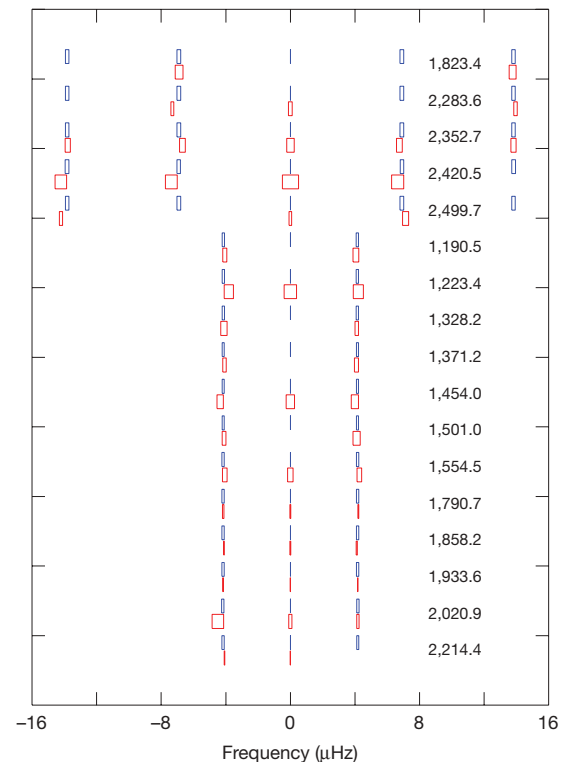


Figure 2 | Observed and calculated rotational splittings in PG 1159–035. This shows the comparison of the observed multiplet structures (in red) with those predicted (in blue) on the basis of the optimal model assuming solid-body rotation with a period of 33.61 ± 0.59 h. The results are expressed in terms of frequency instead of period because, in that case, the predicted multiplet components are equidistant. The uncertainties on the calculated values of the frequency of the components are based on the uncertainty of 0.59 h on the rotation period, while the central component (not affected by rotation in the context of our approach) is assumed to be known to perfect accuracy. The width of each of the blue boxes corresponds to the uncertainty on the calculated frequency of the multiplet component, expressed in units of microhertz. In comparison, the uncertainties on each of the detected (red) frequency components are the actual measured values coming out of the analysis of the multi-site data on PG 1159–035. They depend on the observed amplitudes in the sense that the estimated error on a frequency component is less if its amplitude is large. Not all of the multiplet components have been detected owing to the finite sensitivity of the available observations. This is a common situation encountered in pulsating white dwarfs and reflects the fact that the amplitudes of the components vary significantly from mode to mode. The numbers listed give the central frequency (in microhertz) of each of the 17 multiplets (12 dipole and 5 quadrupole modes). A total of 31 individual spacings has been found in the available data.

relatively large amplitudes deeper inside these stars than in cooler types of pulsating white dwarfs. It will be interesting to see whether other pulsating white dwarfs also exhibit solid rotation in the layers that can be probed with gravity waves. Pressure modes would be the ideal sounding probes of the deeper core in cooler white dwarfs, but, even though pressure mode instabilities are predicted by theory in models of these stars, such oscillations have yet to be found in a real white dwarf^{8,17,18}.

The results obtained in this paper have significant implications for theories of angular momentum transfer in stars^{6,7,19,20}. As a (presumed) typical descendant of a low- to intermediate-mass main-sequence star, PG 1159–035 has been shown to have lost essentially all of its angular momentum and this must be accounted for. In particular, our findings favour the class of theories which imply a very strong coupling between the core and the envelope of stars in terms of angular momentum transport in evolutionary phases before the white-dwarf stage. With such a strong coupling, a star presumably transfers and ultimately loses most of its internal angular momentum during the significant mass-loss episodes associated with the red-giant phases of its evolution. As such, it would enter the white-dwarf stage as a very slow rotator—as we

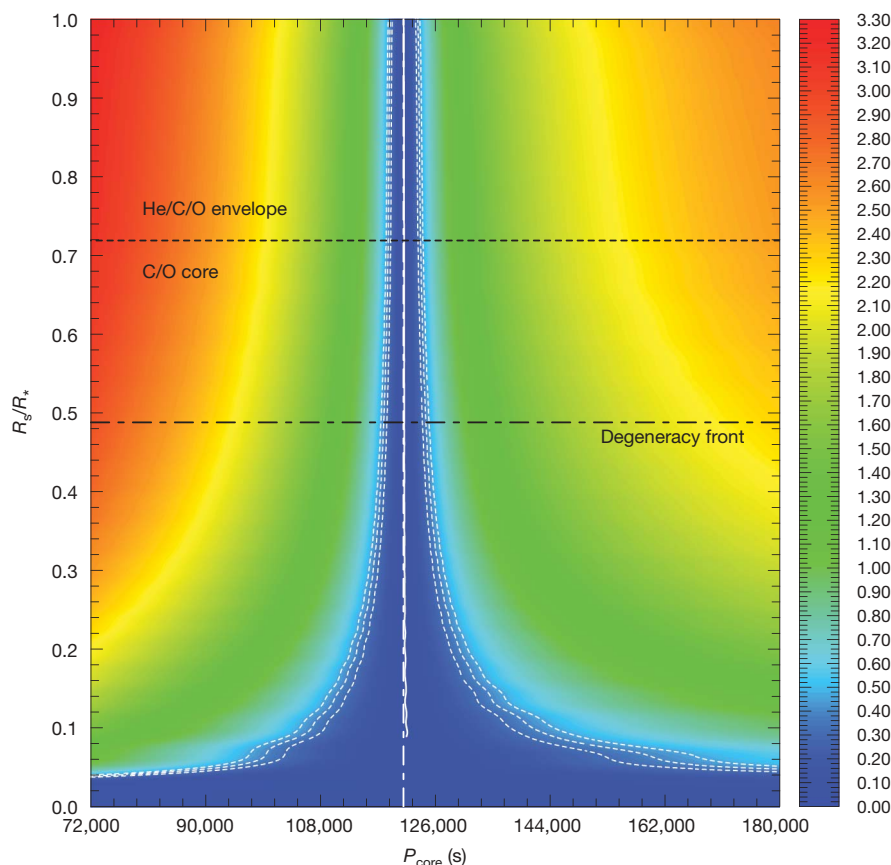


Figure 3 | Internal rotation profile of PG 1159–035. This is a contour map showing the behaviour of the normalized merit function S^2 in terms of depth (expressed in fractional radius; at the surface $R_s/R_* = 1$) and in terms of the rotation period of the inner region in our two-zone approach. We note the logarithmic scale used for S^2 , indicating a very well defined solution. That solution is illustrated by the nearly vertical white curve and the dotted white curves depicting its associated 1, 2 and 3 σ contours. In comparison, the vertical dot-dashed white curve gives the exact solution for solid-body rotation. The fact that the 1, 2 and 3 σ contours diverge out at large depths

indicates that the gravity modes that we use in this experiment lose their capacity at measuring the local rotation rate there. This figure demonstrates explicitly that PG 1159–035 rotates rigidly with a period of 33.61 h over more than 90% of its radius, a region containing more than 97.5% of its mass. The horizontal dashed line is used to indicate the layered chemical structure of the unperturbed seismic model of PG 1159–035. The dot-dashed line depicts the boundary of the degenerate core. The colour scale refers to the logarithm (base 10) of the dimensionless merit function S^2 .

have found here. Clearly, it is important that other studies similar to the one reported here on the pulsating GW Vir star PG 1159–035 be carried out in the future.

Received 3 May; accepted 16 July 2009.

- Fontaine, G. & Wesemael, F. in *Encyclopedia of Astronomy and Astrophysics* 1894–1901 (IOP, 2000).
- Berger, L. *et al.* Rotation velocities of white dwarfs determined from the CaII K line. *Astron. Astrophys.* **444**, 565–571 (2005).
- Koester, D., Dreizler, S., Weidemann, V. & Allard, N. F. Search for rotation in white dwarfs. *Astron. Astrophys.* **338**, 612–622 (1998).
- Ferrario, L. & Wickramasinghe, D. Magnetic fields and rotation in white dwarfs and neutron stars. *Mon. Not. R. Astron. Soc.* **356**, 615–620 (2005).
- Schmidt, G. D. & Smith, P. S. A search for magnetic fields among DA white dwarfs. *Astrophys. J.* **448**, 305–314 (1995).
- Langer, N. Rotation in white dwarfs: Stellar evolution models. *Astron. Soc. Pacif. (Conf. Ser.)* **372**, 3–8 (2007).
- Suijs, M. P. L. *et al.* White dwarf spins from low-mass stellar evolution models. *Astron. Astrophys.* **481**, L87–L90 (2008).
- Fontaine, G. & Brassard, P. The pulsating white dwarf stars. *Publ. Astron. Soc. Pacif.* **120**, 1043–1096 (2008).
- Costa, J. E. S. *et al.* The pulsation modes of the pre-white dwarf PG 1159–035. *Astron. Astrophys.* **477**, 627–640 (2008).
- Córsico, A. H. *et al.* Asteroseismological measurements on PG 1159–035, the prototype of the GW Virginis variable stars. *Astron. Astrophys.* **478**, 869–881 (2008).
- Unno, W. *et al.* *Nonradial Oscillations of Stars* (Tokyo University Press, 1989).
- Charpinet, S. *et al.* Testing the forward modeling approach in asteroseismology. II. Structure and internal dynamics of the hot B subdwarf component in the close eclipsing binary system PG 1336–018. *Astron. Astrophys.* **489**, 377–394 (2008).

- Tassoul, M. & Tassoul, J.-L. Meridional circulation in rotating stars. V. Cooling white dwarfs. *Astrophys. J.* **267**, 334–339 (1983).
- Rabello-Soares, M. C., Basu, S. & Christensen-Dalsgaard, J. On the choice of parameters in solar-structure inversion. *Mon. Not. R. Astron. Soc.* **309**, 35–47 (1999).
- Charbonneau, P. *et al.* Helioseismic constraints on the structure of the solar tachocline. *Astrophys. J.* **527**, 445–460 (1999).
- Di Mauro, M. P. & Dziembowski, W. A. Differential rotation of the solar interior: new helioseismic results by inversion of the SOI-MDI/SOHO data. *Mem. Soc. Astron. Ital.* **69**, 559–562 (1998).
- Saio, H., Winget, D. E. & Robinson, E. L. Pulsation properties of DA white dwarfs—radial mode instabilities. *Astrophys. J.* **265**, 982–995 (1983).
- Silvotti, R. *et al.* Search for p-mode pulsations in DA white dwarfs with VLT-ULTRACAM. *Astron. Soc. Pacif. (Conf. Ser.)* **372**, 593–596 (2007).
- Langer, N., Heger, A., Wellstein, S. & Herwig, F. Mixing and nucleosynthesis in rotating TP-AGB stars. *Astron. Astrophys.* **346**, L37–L40 (1999).
- Spruit, H. C. Dynamo action by differential rotation in a stably stratified stellar interior. *Astron. Astrophys.* **381**, 923–932 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported in part by the NSERC (Canada) and in part by the FQRNT (Québec). G.F. also acknowledges the contribution of the Canada Research Chair Program.

Author Contributions All three authors have contributed equally to this Letter.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to G.F. (fontaine@astro.umontreal.ca).

LETTERS

Violation of Bell's inequality in Josephson phase qubits

Markus Ansmann¹, H. Wang¹, Radoslaw C. Bialczak¹, Max Hofheinz¹, Erik Lucero¹, M. Neeley¹, A. D. O'Connell¹, D. Sank¹, M. Weides¹, J. Wenner¹, A. N. Cleland¹ & John M. Martinis¹

The measurement process plays an awkward role in quantum mechanics, because measurement forces a system to 'choose' between possible outcomes in a fundamentally unpredictable manner. Therefore, hidden classical processes have been considered as possibly predetermining measurement outcomes while preserving their statistical distributions¹. However, a quantitative measure that can distinguish classically determined correlations from stronger quantum correlations exists in the form of the Bell inequalities, measurements of which provide strong experimental evidence that quantum mechanics provides a complete description^{2–4}. Here we demonstrate the violation of a Bell inequality in a solid-state system. We use a pair of Josephson phase qubits^{5–7} acting as spin-1/2 particles, and show that the qubits can be entangled^{8,9} and measured so as to violate the Clauser–Horne–Shimony–Holt (CHSH) version of the Bell inequality¹⁰. We measure a Bell signal of 2.0732 ± 0.0003 , exceeding the maximum amplitude of 2 for a classical system by 244 standard deviations. In the experiment, we deterministically generate the entangled state, and measure both qubits in a single-shot manner, closing the detection loophole¹¹. Because the Bell inequality was designed to test for non-classical behaviour without assuming the applicability of quantum mechanics to the system in question, this experiment provides further strong evidence that a macroscopic electrical circuit is really a quantum system⁷.

In classical physics, deterministic laws provide a complete description for the evolution of a physical system. Quantum physics purports to provide an equally complete description, but the measurement process involves additional premises, and measurement outcomes are intrinsically uncertain. When performing measurements on entangled particles, however, the unpredictability of measurement is combined with very strong correlations between measurements on the individual particles, leading to the apparently paradoxical thought experiments developed by Einstein, Podolsky and Rosen¹².

The CHSH protocol¹⁰ describes one such experiment, with a statistical test to distinguish classical pre-determination from quantum theory. This protocol uses a pair of spin-1/2 particles *A* and *B*, with spin states $|0\rangle$ and $|1\rangle$, which are entangled in the Bell singlet $(|01\rangle - |10\rangle)/\sqrt{2}$. The entangled spins are separated and measured independently along one of two directions (*a* and *a'* for spin *A*, *b* and *b'* for spin *B*). The measurements yield a 0 or 1 for each spin, regardless of the measurement axis. For $x \in a, a'$ and $y \in b, b'$, we define the correlation $E(x, y)$ for the two spins as the difference in the probabilities of measuring the same result versus measuring a different result

$$E(x, y) = P_{\text{same}}(x, y) - P_{\text{diff}}(x, y) \quad (1)$$

$$= P_{00}(x, y) + P_{11}(x, y) - P_{01}(x, y) - P_{10}(x, y)$$

The Bell signal *S* is then defined as

$$S = E(a, b) + E(a', b) - E(a, b') + E(a', b') \quad (2)$$

Classical (predetermined) outcomes result in a Bell signal $|S| \leq 2$, whereas quantum mechanics permits a larger signal $|S| \leq 2\sqrt{2} = 2.828$, for the appropriate measurement axes. Completely random outcomes result in $S = 0$. An experiment returns a Bell violation if $|S| > 2$, and thus indicates quantum entanglement.

The derivation of the limit $|S| \leq 2$ is based on two assumptions, which, if not met, provide loopholes that, in principle, allow an experiment to return a Bell violation even for a classically predetermined process.

The first loophole is called the 'detection loophole'¹¹ and affects experiments in which the spin measurement is ineffective, for example not detecting one of the spins, or confusing a spin from one pair with that of another. This breaks the assumption that a measurement always returns a 0 or 1 for both particles in the pair, because it allows for the additional measurement outcome of 'undetected' or incorrect pair identification. This loophole commonly affects experiments based on the measurement of entangled photons⁴, because a fraction of these are missed by even the best available photon detectors. The reported data set then consists of only a subset of the entire ensemble of entangled photons, a subset whose detection could introduce an unknown classical correlation. This loophole is commonly countered by the 'fair-sampling hypothesis', which claims that no such correlation should exist.

The second loophole, the 'locality/causality loophole', applies when the spin measurements are not performed with true space-like separation, allowing the spins in principle to communicate during the measurement. This loophole affects experiments where the distance *d* between the spins during measurement does not fulfill $d \geq ct_{\text{meas}}$, where t_{meas} specifies the time it takes to completely measure the spins and *c* is the speed of light.

A variety of experiments have shown violations of the Bell inequality^{13–15} with one or the other of these loopholes closed. With the caveat that no one experiment has closed both loopholes, it appears that quantum mechanics provides a more accurate description than do local hidden variable theories.

Here we describe measurements on a pair of Josephson phase qubits, serving as spin-1/2 particles, which are entangled via an electromagnetic resonator^{16,17}. Given that we can generate the entangled pair with certainty and obtain a measurement of each qubit every time, our experiment is not subject to the detection loophole and does not need to invoke the fair-sampling hypothesis. However, the qubits are separated by only 3.1 mm, and with the measurement process lasting around 30 ns our experiment cannot close the locality loophole.

The Josephson phase qubit, as described previously¹⁸, has qubit states $|0\rangle$ and $|1\rangle$ whose energy difference E_{10} can be adjusted by an external current bias. By also applying microwaves at the transition

¹Department of Physics, University of California, Santa Barbara, California 93106, USA.

frequency $E_{10}/h \approx 7$ GHz, the qubit state can be completely controlled. The state is measured by applying a pulse of current, which selectively tunnels the $|1\rangle$ state to an auxiliary state that can be readily distinguished from $|0\rangle$ by a classical measurement of magnetic flux.

In previous experiments two qubits were entangled in a Bell singlet through capacitive coupling¹⁹. This fixed coupling unavoidably keeps the qubits coupled during measurement, so that a measurement of the $|1\rangle$ state of one qubit sometimes produces a $|0\rangle \rightarrow |1\rangle$ transition in the other qubit²⁰. This measurement crosstalk introduces a correlation that complicates the Bell measurement²¹, and in addition inhibits the adjustment of the measurement pulse for optimum fidelity. To circumvent this problem, here we have coupled the two qubits via a resonator, as depicted in Fig. 1. The resonator, with a resonance frequency of 7.185 GHz, acts as a bandpass filter between the two qubits, so that negligible energy is transferred when a qubit $|1\rangle$ is measured. As a result, measurement crosstalk is now found to be unimportant (0.5%, as shown in the Supplementary Information), and the measurement fidelities of 94.6% and 93.4% are within a few per cent of the predicted maximum of 96.6% (ref. 22). The crosstalk reduction was key to the success of this experiment.

Although the transfer of qubit entanglement through the resonator slightly complicates the control sequence, as shown in Fig. 1, the fidelity is not significantly degraded, as demonstrated in previous qubit-resonator experiments^{23,24}. In Fig. 2 we show high-fidelity swapping of an excitation between the qubits, with the entanglement passing through the resonator. Here, the coupling strengths (splittings) of qubits A and B to the resonator are $2g = 26.1$ MHz and 36.2 MHz, respectively; swapping is turned off by detuning the qubits from the resonance frequency of the resonator. We chose different off-detuning frequencies, 530 MHz and 430 MHz for qubits A and B, to minimize crosstalk between each qubit's microwave drive and the other qubit. The qubits are maximally entangled when the probabilities P_{01} and P_{10} first cross at 9.9 ns. Using state tomography¹⁹, we find a fidelity of this state with respect to the Bell singlet $|\psi_s\rangle$ of $F(\rho) = \sqrt{\langle \psi_s | \rho | \psi_s \rangle} = 88.3\%$ and the state's entanglement of formation²⁵ is 0.378.

This entangled state is used for the Bell violation experiment. To measure the Bell state along different axes (a, a') and (b, b'), we first rotate the qubit states using microwave pulses before performing

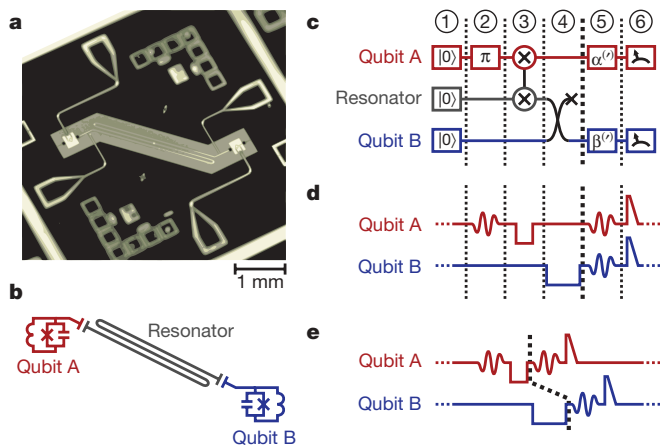


Figure 1 | Resonator-coupled qubits. **a**, Photograph of qubit die showing two qubits coupled via a coplanar waveguide resonator. **b**, Circuit diagram of **a**. **c**, Experimental sequence in quantum circuit notation. (1) Initialize to $|00\rangle$ state. (2) Create state $|10\rangle$ via π -pulse to qubit A. (3) Entangle qubit A with resonator via square-root of i -SWAP coupling¹⁹. (4) Resonator entanglement swapped to qubit B, creating a generalized Bell state $(|10\rangle - e^{i\theta}|01\rangle)/\sqrt{2}$. (5) Rotate state for change of measurement axes. (6) Tunnelling measurement (along z axis) of qubit states. **d**, Control sequence used to implement **c**. Downward square pulses represent tuning the qubits into resonance with the resonator ('3' and '4'), whereas oscillations represent microwave excitations ('2' and '5'); triangular pulses at the end are for measurement ('6'). **e**, Shortened control sequence. The removal of dead time places measurement of qubit A about 11 ns before B.

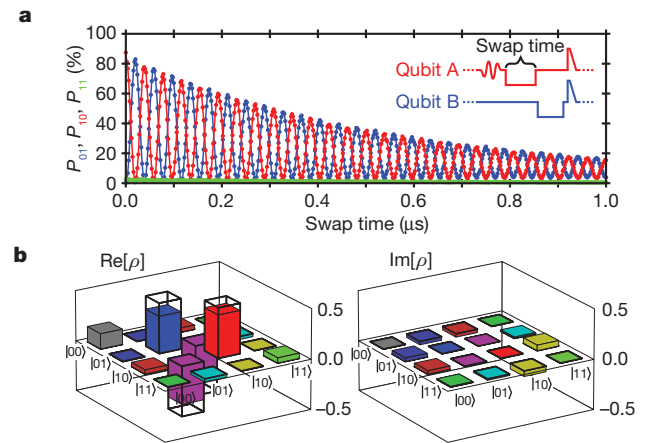


Figure 2 | Entanglement analysis. **a**, Measurement probabilities of $|01\rangle$ (blue), $|10\rangle$ (red), and $|11\rangle$ (green) plotted as a function of swap time between qubit A and the resonator. The inset shows the experimental sequence, where the length of the coupling pulse for qubit B is set to produce a state-swap operation between the resonator and qubit. The qubits are maximally entangled at the first crossing of P_{01} and P_{10} , at an interaction time of 9.9 ns. **b**, Measurement of the real and imaginary part of the density matrix ρ for this entangled state using state tomography¹⁹, after including an azimuthal (z -axis) rotation of 18° to compensate for the accumulated relative phase θ due to coupling pulses and differing qubit frequencies. The fidelity of our entangled state with respect to the ideal Bell singlet $|\psi_s\rangle$ (box outlines) is $F(\rho) = \sqrt{\langle \psi_s | \rho | \psi_s \rangle} = 88.3\%$. To understand better the actual state of the entangled pair that will undergo the Bell measurement, the density matrix can be corrected for measurement errors to give a state fidelity of 92.1%. The state's entanglement of formation²⁵ is 0.378 (0.449 corrected).

measurements along the z axis of the Bloch sphere¹⁹, as depicted in Fig. 1c and d. Because the coherence of the entangled state degrades in time due to energy decay and dephasing, we additionally shorten the pulse sequence as much as possible, as depicted in Fig. 1e.

The CHSH version of the Bell inequality is not based on any assumptions about the entangled state or the choice of measurement axes, so we use search optimization of all relevant parameters in the sequence to maximize S . We find that this search always converges to a violation with $|S| > 2$, using sequence parameters that make physical sense (see Supplementary Information). For example, the measurement axes are close to those expected for maximum violation, with the angle between a and a' (or b and b') close to 90° , and the relative in-plane angle between a' and b' close to 45° . However the plane of (a, a') is rotated by an arbitrary azimuthal angle from (b, b') as a phase shift between the states $|01\rangle$ and $|10\rangle$ is produced by the differing qubit frequencies and the tuning pulses that bring the qubits on resonance with the resonator (see Fig. 3).

With optimal parameters, we measure a Bell signal with $S = 2.0732 \pm 0.0003$, which corresponds to a violation by 244 standard deviations. This value is obtained from an average over 34.1 million runs of the sequence. We estimate that with perfect measurement fidelities the Bell signal would be $S = 2.355$ (see Supplementary Information).

Given that this result is only a single number, it is important to perform verification experiments to check for errors. After all, turning off the measurement electronics gives $P_{00} = 1$ and $S = 2$. In Fig. 3a, we plot S as a function of the azimuthal angle between the (a, a') and (b, b') measurement planes, or, equivalently, the phase between the microwave rotation pulses applied to qubits A and B. The sinusoidal dependence is as expected from theory. In Fig. 3b we plot S versus the time delay between the two measurement pulses: We find that S decays with time as expected from the loss of qubit coherence, with no discernible effect when the measurement pulses overlap²⁰, as expected for negligible crosstalk. Finally, we compare the measured value of S with that expected from theory. Quantum simulations (see Supplementary Information) predict that S is reduced from its theoretical maximum of $2\sqrt{2} = 2.828$ to $S = 2.500$, owing to the finite energy relaxation times

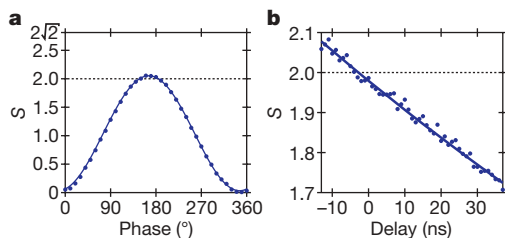


Figure 3 | Verification experiments. Points are experimental data, and lines are theory. **a**, Plot of S versus the phase difference between the microwave pulses applied to qubits A and B . This is equivalent to varying the azimuthal (z -axis) angle between the plane containing the measurement axes a and a' with respect to the (b, b') plane. Sinusoidal modulation is observed, as expected. The phase-offset of the oscillation results from accumulated phases between the qubits due to coupling pulses and differing qubit frequencies. **b**, Plot of S versus relative time delay between the measurement pulses applied to qubits A and B . A reduction in S versus delay is observed, consistent with the loss of qubit coherence.

T_1 that were measured to be 392 ns, 296 ns and 2,550 ns for qubit A , qubit B and the resonator, respectively. The value of S is reduced further by dephasing to $S = 2.337$ using the measured dephasing times $T_2 = 146$ ns, 135 ns and $\sim 5,000$ ns. Finally, the measurement fidelities of 94.6% and 93.4% give $S = 2.064$, within 0.5% of our measured value. The slightly larger measured S is probably due to our dephasing calculation, which assumes uncorrelated (white) flux noise²⁶ that slightly overestimates the effects of dephasing, and our maximization of S accounts for small asymmetries in qubit parameters, which are not included in the simulations.

Another comparison of fidelity is shown in Fig. 4, in which we display data and simulations for the qubit–resonator–qubit swap outcome. We observe excellent agreement using only the measured single qubit error mechanisms of decoherence and measurement fidelity. With the difference between experiment and theory after four full swap operations (arrow 2 in Fig. 4) being less than 1%, we believe the errors due to the entanglement operation, corresponding to two-qubit errors, are less than 1%. This is not unexpected as prior entangling experiments between a qubit and a resonator gave high fidelities^{23,24}, and the capacitor coupling elements and resonator used here have low dissipation.

In conclusion, we have measured a violation of the CHSH Bell inequality in a macroscopic solid-state quantum system that closes the detection loophole. We note that successfully performing this measurement requires the simultaneous optimization of a number of qubit performance benchmarks, which together form most of what are known as the DiVincenzo criteria for a quantum computational architecture²⁷. The very high demands on qubit state preparation, entanglement and measurement required to achieve a violation of a

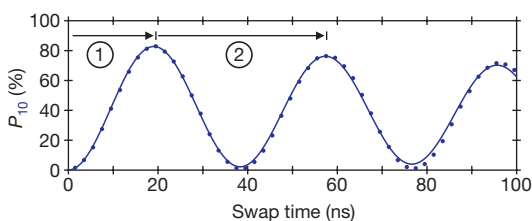


Figure 4 | Entanglement analysis. Probability of measuring the $|01\rangle$ final qubit state versus the qubit A –resonator swap time (dots) along with results of quantum simulations (line), as in Fig. 2a. Arrow ‘1’ indicates the time for a qubit–resonator swap, while arrow ‘2’ is the time for three full swap operations; both ‘1’ and ‘2’ are followed by a resonator–qubit B swap before measurement (not shown). Quantum simulations (see Supplementary Information), which only account for single qubit coherence and measurement fidelity, match the data with excellent agreement even after four full swaps (‘2’). The increasing disagreement after ~ 70 ns is caused by finite settling times in the flux bias circuitry, making qubit A slowly drift off-resonance from the resonator. We conclude that errors due to the entanglement operation, corresponding to two-qubit errors, are less than 1%.

Bell inequality make this type of measurement useful as a single-number benchmark for any quantum computation implementation, forming the basis for comparing very different physical architectures. We also note that the fast and independent measurement of the entangled qubits should enable, at least in principle, closure of the locality loophole in future extensions of this experiment.

Received 25 May; accepted 4 August 2009.

1. Bell, J. On the Einstein-Podolsky-Rosen paradox. *Physics* **1**, 195–200 (1964).
2. Roos, C. F. *et al.* Bell states of atoms with ultralong lifetimes and their tomographic state analysis. *Phys. Rev. Lett.* **92**, 220402 (2004).
3. Kocher, C. A. & Commins, E. D. Polarization correlation of photons emitted in an atomic cascade. *Phys. Rev. Lett.* **18**, 575 (1967).
4. Aspect, A. *et al.* Experimental tests of realistic local theories via Bell's theorem. *Phys. Rev. Lett.* **47**, 460–463 (1981).
5. Devoret, M. H. & Martinis, J. M. Implementing qubits with superconducting integrated circuits. *Quantum Inf. Process.* **3**, 163–203 (2004).
6. Martinis, J. Superconducting phase qubits. *Quant. Inf. Process.* **8**, 81–103 (2009).
7. Clarke, J. & Wilhelm, F. K. Superconducting quantum bits. *Nature* **453**, 1031–1042 (2008).
8. DiCarlo, L. *et al.* Demonstration of two-qubit algorithms with a superconducting quantum processor. *Nature* **460**, 240–244 (2009).
9. Wei, L. F., Liu, Y. X., Storcz, M. J. & Nori, F. Macroscopic Einstein-Podolsky-Rosen pairs in superconducting circuits. *Phys. Rev. A* **73**, 052307 (2006).
10. Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**, 880–884 (1969).
11. Pearle, P. M. Hidden-variable example based upon data rejection. *Phys. Rev. D* **2**, 1418–1425 (1970).
12. Einstein, A., Podolsky, B. & Rosen, N. Can the quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47**, 777–780 (1935).
13. Weihs, G., Jennewein, T., Simon, C., Weinfurter, H. & Zeilinger, A. Violation of Bell's inequality under strict Einstein locality conditions. *Phys. Rev. Lett.* **81**, 5039–5043 (1998).
14. Rowe, M. A. *et al.* Experimental violation of a Bell's inequality with efficient detection. *Nature* **409**, 791–794 (2001).
15. Matsukevich, D. N., Maunz, P., Moehring, D. L., Olmschen, S. & Monroe, C. Bell inequality violation with two remote atomic qubits. *Phys. Rev. Lett.* **100**, 150404 (2008).
16. Sillanpaa, M. A., Park, J. I. & Simmonds, R. W. Coherent quantum state storage and transfer between two phase qubits via a resonant cavity. *Nature* **449**, 438–442 (2007).
17. Majer, J. *et al.* Coupling superconducting qubits via a cavity bus. *Nature* **449**, 443–447 (2007).
18. Lucero, E. *et al.* High-fidelity gates in a Josephson qubit. *Phys. Rev. Lett.* **100**, 247001 (2008).
19. Steffen, M. *et al.* Measurement of the entanglement of two superconducting qubits via state tomography. *Science* **313**, 1423–1425 (2006).
20. McDermott, R. *et al.* Simultaneous state measurement of coupled Josephson phase qubits. *Science* **307**, 1299–1302 (2005).
21. Kofman, A. G. & Korotkov, A. N. Analysis of Bell inequality violation in superconducting phase qubits. *Phys. Rev. B* **77**, 104502 (2008).
22. Cooper, K. B. *et al.* Observation of quantum oscillations between a Josephson phase qubit and a microscopic resonator using fast readout. *Phys. Rev. Lett.* **93**, 180401 (2004).
23. Hofheinz, M. *et al.* Synthesizing arbitrary quantum states in a superconducting resonator. *Nature* **459**, 546–549 (2004).
24. Hofheinz, M. *et al.* Generation of Fock states in a superconducting quantum circuit. *Nature* **454**, 310–314 (2008).
25. Hill, S. & Wootters, W. K. Entanglement of a pair of quantum bits. *Phys. Rev. Lett.* **78**, 5022–5025 (1997).
26. Bialczak, R. C. *et al.* $1/f$ flux noise in Josephson phase qubits. *Phys. Rev. Lett.* **99**, 187006 (2007).
27. DiVincenzo, D. P. The physical implementation of quantum computation. Preprint at (<http://arxiv.org/abs/quant-ph/0002077>) (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Korotkov and A. Kofman for discussions of our measurement process. Devices were made at the UCSB Nanofabrication Facility, a part of the NSF-funded National Nanotechnology Infrastructure Network. This work was supported by IARPA under grant W911NF-04-1-0204 and by the NSF under grant CCF-0507227.

Author Contributions M.A. performed the experiment and analysed the data, while H.W. fabricated the sample. J.M.M. and E.L. designed the custom electronics and M.H. developed the calibrations for it. M.A. and M.N. provided software infrastructure. All authors contributed to various tasks, such as the fabrication process, qubit design, or experimental set-up.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.M.M. (martinis@physics.ucsb.edu).

Stable isotope constraints on Holocene carbon cycle changes from an Antarctic ice core

Joachim Elsig^{1,2}, Jochen Schmitt^{1,2,3}, Daiana Leuenberger^{1,2}, Robert Schneider^{1,2}, Marc Eyer^{1,2}, Markus Leuenberger^{1,2}, Fortunat Joos^{1,2}, Hubertus Fischer^{1,2,3} & Thomas F. Stocker^{1,2}

Reconstructions of atmospheric CO₂ concentrations based on Antarctic ice cores^{1,2} reveal significant changes during the Holocene epoch, but the processes responsible for these changes in CO₂ concentrations have not been unambiguously identified. Distinct characteristics in the carbon isotope signatures of the major carbon reservoirs (ocean, biosphere, sediments and atmosphere) constrain variations in the CO₂ fluxes between those reservoirs. Here we present a highly resolved atmospheric $\delta^{13}\text{C}$ record for the past 11,000 years from measurements on atmospheric CO₂ trapped in an Antarctic ice core. From mass-balance inverse model calculations^{3,4} performed with a simplified carbon cycle model, we show that the decrease in atmospheric CO₂ of about 5 parts per million by volume (p.p.m.v.) The increase in $\delta^{13}\text{C}$ of about 0.25‰ during the early Holocene is most probably the result of a combination of carbon uptake of about 290 gigatonnes of carbon by the land biosphere and carbon release from the ocean in response to carbonate compensation of the terrestrial uptake during the termination of the last ice age. The 20 p.p.m.v. increase of atmospheric CO₂ and the small decrease in $\delta^{13}\text{C}$ of about 0.05‰ during the later Holocene can mostly be explained by contributions from carbonate compensation of earlier land-biosphere uptake and coral reef formation, with only a minor contribution from a small decrease of the land-biosphere carbon inventory.

The Holocene is the current interglacial period, starting about 11,000 years before present (11 kyr BP, where present is defined as AD 1950) following the Transition (here defined as 18–11 kyr BP) from the last glacial maximum. Variations in the atmospheric concentration of CO₂ during the Holocene were significant but small compared to glacial-interglacial changes of typically 100 p.p.m.v. (refs 5, 6). Yet a decrease of about 5 p.p.m.v. from 11–7.5 kyr BP could be observed, followed by an increase of about 20 p.p.m.v. to the pre-industrial level of about 280 p.p.m.v. (refs 1, 2, 7). Different explanations for these variations were discussed^{7,8}, such as changes in the carbon inventories of vegetation, soils and peatlands⁹, in anthropogenic land use^{10,11}, in sea surface temperature (SST)^{7,12}, coral reef growth^{13,14} or carbonate compensation¹⁵. The latter is a multi-millennial equilibration process of the atmosphere–ocean–sediment system and the weathering cycle. Moreover, model simulations of atmospheric CO₂ and $\delta^{13}\text{C}$ during the Holocene have not provided an unambiguous quantitative explanation^{7,8,16}. The major stumbling block has been the scarcity of reconstructions of $\delta^{13}\text{C}$ on atmospheric CO₂ with sufficient accuracy and time resolution from the last glacial maximum to the Holocene^{7,17–20}.

We performed carbon isotope measurements on air trapped in the EPICA (European Project for Ice Coring in Antarctica) Dome C (75° 06' S, 123° 24' E) ice core using two completely independent extraction methods (mechanical cracking and sublimation, see Supplementary Information). Altogether, 199 single samples have

been measured from 59 different depths in the interval of 110 to 410 m. This interval corresponds to a gas age range of 11–0.35 kyr BP (ref. 21). The results are gravitationally corrected (see Supplementary Information) and are presented in Fig. 1. Of the 165 samples that were extracted with a mechanical cracker, a minimum of two were taken at each depth. For the remaining 34 samples we used a sublimation technique, and these are either single measurements or replications of three adjacent ice samples. The overall precision for a single measurement is 0.07‰ for both methods. The results of the two different methods agree very well within their uncertainties. The record clearly shows a continuous increase in the $\delta^{13}\text{C}$ values during the first 5 kyr of the Holocene, followed by only slightly decreasing values.

We focus on the evolution of the carbon isotopes on a timescale of a few thousand years. Therefore, we calculated a spline and its 1 σ uncertainty bands with a cut-off period of 5 kyr (Fig. 2). In a Monte Carlo simulation, standard deviations smaller than 0.07‰ were

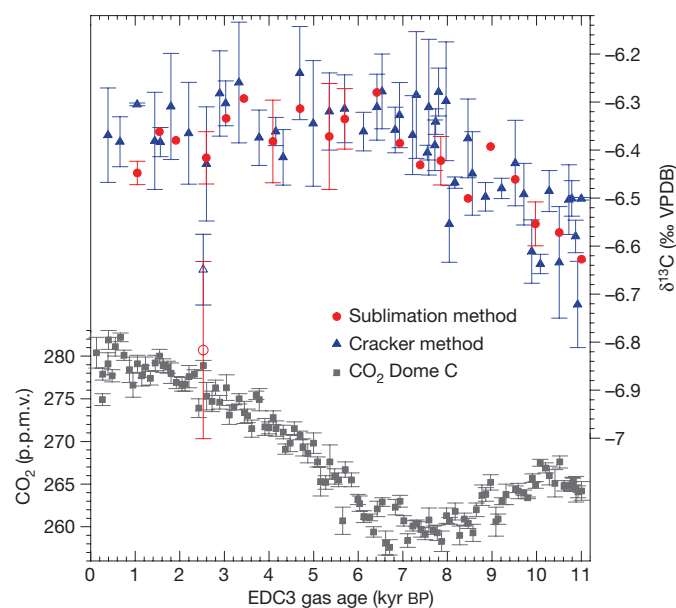


Figure 1 | $\delta^{13}\text{C}$ and CO₂^{1,2} measured in air trapped in ice from Dome C, Antarctica. Blue triangles indicate measurements performed with the cracker (mean of two to four samples); red circles indicate measurements with the sublimation method (single measurements or mean of three adjacent samples). Open symbols indicate outliers. The error bars represent the t -weighted 1 σ standard deviations of the mean. Grey squares represent CO₂ data from Dome C (mean of six samples; error bars, 1 σ of the mean)²¹. EDC3, EPICA Dome C timescale version 3.

¹Climate and Environmental Physics, Physics Institute, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland. ²Oeschger Centre for Climate Change Research, University of Bern, Zähringerstrasse 25, CH-3012 Bern, Switzerland. ³Alfred Wegener Institute for Polar and Marine Research (AWI), Columbusstrasse, D-27568 Bremerhaven, Germany.

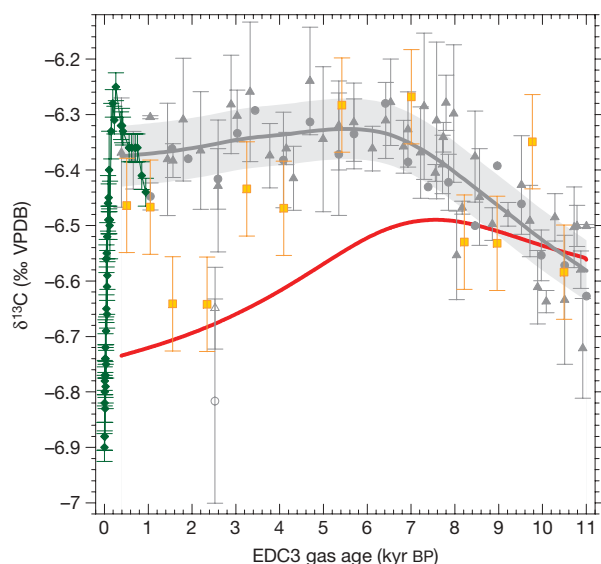


Figure 2 | $\delta^{13}\text{C}$ ice-core records measured on the Antarctic ice cores from Dome C, Taylor Dome⁷ and Law Dome²². Grey circles (sublimation method, this work) and grey triangles (cracker method, this work) refer to results from Dome C, yellow squares are data from Taylor Dome, and green diamonds are data from Law Dome. Grey open symbols indicate outliers. The error bars for Dome C measurements are the same as in Fig. 1. The error bars for the Taylor Dome and Law Dome data represent 1σ of the mean. The grey spline through the Dome C data represents the mean of one hundred Monte Carlo simulations with a cut-off period of 5 kyr. The grey shaded area indicates $\pm 1\sigma$ standard deviation of the spline. The red line is the result of a deconvolution of the atmospheric CO_2 record, assuming the land-biosphere only scenario. VPDB, Vienna Pee-Dee Belemnite standard.

increased to 0.07‰; those higher than 0.07‰, however, were retained. Measurements at one depth interval (open symbols at 2.5 kyr BP in Fig. 1) led to exceptionally negative $\delta^{13}\text{C}$ values significantly outside the 2σ uncertainty range independently for both extraction methods. The reason for these outliers remains obscure, but the very large scatter in neighbouring samples despite the low-pass filtering effect of the bubble-enclosure process rules out an atmospheric origin of these outliers (see Supplementary Information). Accordingly, these two outliers are not included in the calculation of the spline.

Two main features of the carbon isotope record can be recognized: first, an almost linear increase from -6.58‰ to -6.33‰ between 11 and 6 kyr BP, and second, a small decrease of 0.05‰ in the later Holocene after 6 kyr BP. In Fig. 2 our data are compared with two published $\delta^{13}\text{C}$ records, from the Taylor Dome⁷ and Law Dome²² ice cores. The relative timing of the Taylor Dome to Dome C gas age was adapted from Monnin *et al.*¹, who matched the CO_2 records of Dome C and Taylor Dome for the Holocene. The Taylor Dome data generally agree with our measurements except for three data points (out of 12) that have no overlap within the reported 1σ uncertainties of our record. Our data are entirely consistent with the Law Dome record covering the last thousand years, providing a direct link to atmospheric measurements through the firn gas data from Law Dome²².

We used our new, more precise and better resolved $\delta^{13}\text{C}$ record to quantify carbon fluxes with different isotopic signatures and to test various hypotheses of Holocene CO_2 variations. Previous mechanistic model studies^{8,14,16,23}, ocean sediment analyses¹⁵, coral reef reconstructions²⁴, and peatland data^{9,25} suggest that the Holocene CO_2 variations were caused by land-biosphere uptake, carbonate compensation of earlier land-biosphere uptake, and the build-up of coral reefs made of calcium carbonate. In contrast, changes in SST¹², ocean circulation and marine biological cycling, shifts in the proportion of C_3 to C_4 photosynthesis⁸, or changes in volcanic outgassing are assumed to be

of minor importance for the Holocene CO_2 and $\delta^{13}\text{C}$ evolution. In addition, a strong release of CO_2 over the past 7,000 years from anthropogenic land-use changes has previously been proposed¹⁰. The multitude and the spatio-temporal variability of the processes influencing atmospheric CO_2 and $\delta^{13}\text{C}$ prevent us from firmly attributing the measured changes to a single mechanism. However, the deconvolution of our new record by mass-balance inverse calculations^{3,4,7} (see Supplementary Information) permits us to quantify atmosphere–ocean and atmosphere–land carbon fluxes for different hypotheses and to attribute the measured CO_2 and $\delta^{13}\text{C}$ changes to the most likely mechanisms.

Model calculations were performed with a cost-efficient impulse response representation of carbonate compensation and of the High-Latitude Exchange/Interior Diffusion–Advection ocean model coupled to a 4-box representation of vegetation and soils and a well-mixed atmosphere. The substitute model yields results for atmospheric CO_2 and the redistribution of carbon and carbon isotopes between global reservoirs that are comparable to those of spatially resolved models. For instance, for the main scenario of land carbon uptake discussed below, the substitute model yields a late Holocene CO_2 rise of 15 p.p.m.v. compared to 12 p.p.m.v. obtained with the Bern3D dynamic ocean–sediment model.

First, we assessed a land-biosphere-only scenario. Atmospheric CO_2 variations were assumed to be driven entirely by changes in the land biosphere (using a fractionation for photosynthesis of 18.7‰; ref. 3) and the ocean–sediment system was assumed to react only passively. Solving the atmospheric CO_2 budget for the unknown terrestrial flux yields a land-biosphere uptake of 75 gigatonnes of carbon (Gt C) from 11 to 7 kyr BP and a release of 275 Gt C thereafter⁷. More importantly, the simulated evolution of atmospheric $\delta^{13}\text{C}$ is not compatible with the $\delta^{13}\text{C}$ measurements (Fig. 2). The release of isotopically depleted terrestrial carbon yields a modelled $\delta^{13}\text{C}$ decrease of 0.25‰ after 8 kyr BP, whereas our data show little change. Thus, suggestions that CO_2 emissions from anthropogenic land-use change caused the late Holocene CO_2 rise¹⁰ are quantitatively inconsistent with our $\delta^{13}\text{C}$ record, as well as with other evidence⁸.

Other scenarios considering only one driving mechanism (changes in SST, the marine biological cycle, or the calcium carbonate cycle only), or the combination of marine biological changes and land-biosphere changes, are also in conflict with the measured co-evolution of CO_2 and $\delta^{13}\text{C}$ and with other proxy evidence. For example, an unrealistically large global average SST increase of 2.5 °C is required to explain the $\delta^{13}\text{C}$ increase of 0.25‰ from 11 to 6 kyr BP by SST changes only. A combination of SST and land-biosphere changes (tested by deconvolving both atmospheric records simultaneously) implies a global average SST increase of about 1.5 °C, which is in conflict with alkenone-based SST reconstructions (0.2 ± 0.2 °C since 8 kyr BP, ref. 26) and model calculations^{12,27}. This is different from earlier findings by Indermühle *et al.*⁷, who applied the same modelling approach but had to rely on very few $\delta^{13}\text{C}$ values. They suggested that changes in the land biosphere with a modest temperature increase, possibly in combination with changes in the marine-carbonate cycle, are responsible for the Holocene CO_2 evolution. Our improved $\delta^{13}\text{C}$ record permits us now to exclude this land-biosphere–global SST scenario.

Finally, we turn to the land-biosphere–marine-carbonate scenario, which assumes that atmospheric variations are caused by changes in the land biosphere in combination with carbonate compensation and coral reef growth. Both marine-carbonate processes are associated with small isotopic fractionations and cannot be distinguished by $\delta^{13}\text{C}$ data, nor can they significantly influence atmospheric $\delta^{13}\text{C}$. Solving the two atmospheric budgets for CO_2 and $\delta^{13}\text{C}$ yields a land-biosphere uptake of (290 ± 36) Gt C (mean and $\pm 1\sigma$ confidence interval from a Monte Carlo analysis) during the period 11–5 kyr BP and a release of (36 ± 37) Gt C thereafter (Fig. 3a). We note that the substitute model represents global mean fluxes and is not able to capture the influence of spatial variations in carbon and isotopic

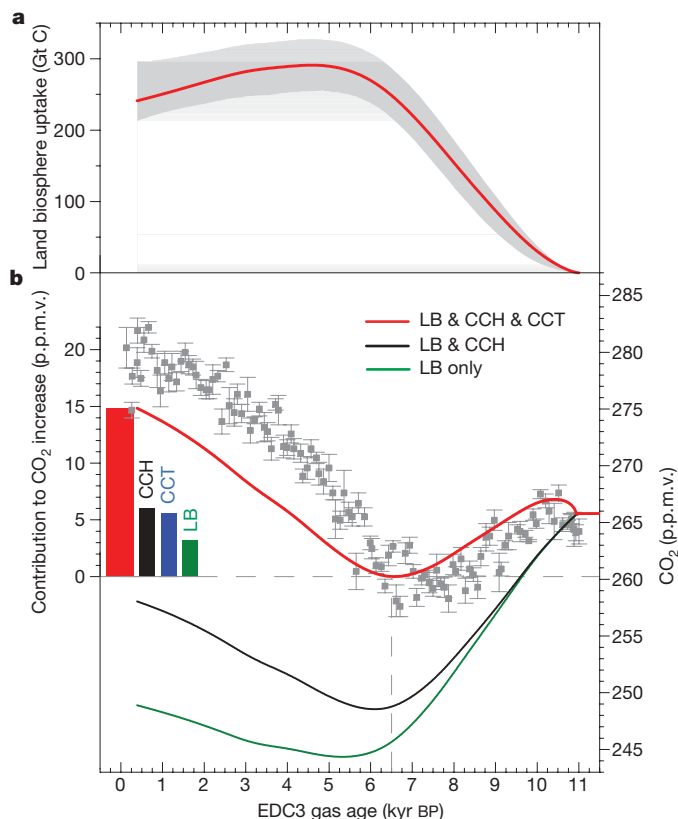


Figure 3 | Attribution of simulated CO₂ to different processes.

a, Cumulative land-biosphere uptake for the land-biosphere–marine-carbonate scenario from deconvolving the atmospheric CO₂ (refs 1, 2) and $\delta^{13}\text{C}$ records. The shaded area indicates the 1 σ confidence interval of a Monte Carlo analysis⁴, taking into account the uncertainty of the ice-core data. **b**, Atmospheric CO₂ is simulated by prescribing the land-biosphere changes shown in **a**. Grey squares represent CO₂ data from Dome C (mean of six samples; error bars, 1 σ of the mean). Green line, land biosphere (LB) only; black line, LB and carbonate compensation during the Holocene (CCH); red line, LB plus CCH plus 700 Gt C land biosphere uptake during the transition which causes carbonate compensation (CCT). The bar chart indicates the modelled contributions to the CO₂ rise from 6.5–0 kyr BP by individual processes LB, CCT and CCH, and the total (red bar), assuming the land-biosphere–marine-carbonate scenario. The remaining difference between the simulated and measured CO₂ increase may be attributed to coral reef growth and other mechanisms.

exchange. Potential contributions from other processes such as volcanism, SST changes or changes in the marine biological cycle are neglected in the land-biosphere–marine-carbonate scenario.

Are these fluxes inferred from our mass-balance calculation realistic and in agreement with proxy data, model results and process understanding? The early Holocene land-biosphere uptake of 290 Gt C is compatible with an early Holocene uptake of 110 Gt C simulated by the Lund–Potsdam–Jena Dynamic Vegetation model on non-peatland land in response to reconstructed ice-sheet retreat, climate and CO₂ variations⁸. This requires an early Holocene increase in peat carbon by 180 Gt C. Current inventories of peat carbon range from 270 to 455 Gt C (ref. 9). Radiocarbon dating suggests initiation of peat accumulation already occurring during the Transition with a major expansion of peat area in the early Holocene⁹, but also some peat accumulation in recent millennia²⁸. A land-biosphere release of ~40 Gt C in the later Holocene may be explained by a model-estimated release of 30 Gt C due to the desertification of the Sahara⁷, plus a modelled release of around 50 Gt C from pre-1,500 AD anthropogenic land-use change¹¹, plus an implied peat carbon uptake of ~40 Gt C. The implied Holocene peat carbon uptake of 220 Gt C is somewhat lower than the published range for the current carbon inventory in peat carbon⁹. This might be explained by peat carbon uptake during the

Transition. Alternatively, natural carbon release from land might have been larger than 30 Gt C (ref. 23), also implying a higher uptake by peat.

In a next step, the atmosphere–ocean component of our carbon cycle model was used in forward simulations to test mechanistic explanations of inferred ocean-driven CO₂ changes. First, we prescribe the atmosphere–land fluxes from the land-biosphere–carbonate-compensation scenario to separate the contribution from Holocene land-biosphere changes. The atmospheric $\delta^{13}\text{C}$ record is matched by design. The Holocene land-biosphere changes result in an initial decrease in atmospheric CO₂ of about 17 p.p.m.v., which is much larger than that measured in ice cores, and an increase of ~9 p.p.m.v. after 6.5 kyr BP (Fig. 3b). The increase is attributed to carbonate compensation of the early Holocene land-biosphere uptake (~6 p.p.m.v.) and to the (prescribed) land-biosphere release of 50 Gt C (~3 p.p.m.v.). Second, we consider that carbonate compensation caused by the land-biosphere uptake during the Transition also contributed to the Holocene CO₂ rise. Land-biosphere carbon inventory changes over the glacial–interglacial period have been estimated to be from 500 Gt C to more than 1,000 Gt C (refs 29, 30). In our simulations we have prescribed a linear increase of the land-biosphere inventory by 700 Gt C during the Transition. This yields a simulated early Holocene CO₂ decrease comparable to the ice-core data and a late Holocene rise of ~15 p.p.m.v., somewhat smaller than reconstructed (see Supplementary Information for additional sensitivity simulations).

The remaining ~5 p.p.m.v. in atmospheric CO₂ can plausibly be explained by coral reef growth. Coral reef formation changes the carbonate ion balance in sea water and increases atmospheric CO₂ (ref. 13). Vecsei and Berger²⁴ reconstructed a calcium carbonate accumulation by reefs of 260 Gt C during the past 8 kyr. This implies a CO₂ rise by another ~8 p.p.m.v. (ref. 14), compatible with the ice-core data and model results. However, our modelling results and the $\delta^{13}\text{C}$ record do not quantitatively support the suggestion that coral reef growth caused a much larger CO₂ rise of 20 to 40 p.p.m.v. (ref. 14) during the late Holocene.

High-resolution and high-precision ice-core data on atmospheric $\delta^{13}\text{C}$ can significantly constrain the possible pathways of the carbon cycle evolution in the past. Using carbon cycle modelling, our new $\delta^{13}\text{C}$ data show that the evolution of atmospheric CO₂ over the past 11 kyr is dominated by an early Holocene increase in land biosphere and changes in the marine-carbonate cycle. Our $\delta^{13}\text{C}$ record renders untenable suggestions that CO₂ emissions from anthropogenic land-use changes caused the later Holocene CO₂ rise¹⁰ and prevented a new ice age. The attribution of the CO₂ changes to specific changes in the marine-carbonate cycle (including coral reef growth, carbonate compensation of land-biosphere carbon uptake, sediment–ocean interactions related to the reorganization of the marine carbon cycle during the glacial–interglacial transition, or changes in weathering fluxes) is challenging. Further progress requires an extension of our atmospheric $\delta^{13}\text{C}$ record into the glacial period as well as process studies with three-dimensional physical-biogeochemical climate models.

METHODS SUMMARY

For the cracker method we use a mechanical extraction device (steel needle cracker), which crushes the ice sample (6 g) under vacuum to release the enclosed gases in the ice. In the sublimation method the gases are extracted from the ice sample (30 g) using sublimation at –25 °C in a glass vessel. In both methods the released air is dried in a water trap and the amount of air is measured with a pressure gauge to calculate the CO₂ concentration. CO₂ and N₂O are separated from the major air components at –196 °C using liquid nitrogen with subsequent concentration by means of a cryofocus capillary to ensure complete gas chromatographic separation. After having passed through a chromatographic column to separate isobaric components (N₂O and organic components from drilling fluid) the purified CO₂ sample is injected via an open-split tube into the isotope ratio mass spectrometer (a Delta Plus XL for the cracker method and a MAT 253 for the sublimation method, both ThermoFisher). Both extraction systems are equipped with inlet devices which allow the processing of

calibrated reference gases in the same way as the ice samples. This allows us to check the system and reference the results from the ice samples to an international standard.

Received 4 May; accepted 28 July 2009.

- Monnin, E. *et al.* Atmospheric CO₂ concentrations over the last glacial termination. *Science* **291**, 112–114 (2001).
- Flückiger, J. *et al.* High-resolution Holocene N₂O ice core record and its relationship with CH₄ and CO₂. *Glob. Biogeochem. Cycles* **16**, 1010, doi:10.1029/2001GB001417 (2002).
- Joos, F. & Bruno, M. Long-term variability of the terrestrial and oceanic carbon sinks and the budgets of the carbon isotopes ¹³C and ¹⁴C. *Glob. Biogeochem. Cycles* **12**, 277–295 (1998).
- Bruno, M. & Joos, F. Terrestrial carbon storage during the past 200 years: a Monte Carlo analysis of CO₂ data from ice core and atmospheric measurements. *Glob. Biogeochem. Cycles* **11**, 111–124 (1997).
- Lüthi, D. *et al.* High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature* **453**, 379–382 (2008).
- Sigman, D. M. & Boyle, E. A. Glacial/interglacial variations in atmospheric carbon dioxide. *Nature* **407**, 859–869 (2000).
- Indermühle, A. *et al.* Holocene carbon-cycle dynamics based on CO₂ trapped in ice at Taylor Dome, Antarctica. *Nature* **398**, 121–126 (1999).
- Joos, F., Gerber, S., Prentice, I. C., Otto-Bliesner, B. L. & Valdes, P. J. Transient simulations of Holocene atmospheric carbon dioxide and terrestrial carbon since the Last Glacial Maximum. *Glob. Biogeochem. Cycles* **18**, GB2002, doi:10.1029/2003GB002156 (2004).
- MacDonald, G. M. *et al.* Rapid early development of circumarctic peatlands and atmospheric CH₄ and CO₂ variations. *Science* **314**, 285–288 (2006).
- Ruddiman, W. F. The anthropogenic greenhouse era began thousands of years ago. *Clim. Change* **61**, 261–293 (2003).
- Strassmann, K. M., Joos, F. & Fischer, G. Simulating effects of land use changes on carbon fluxes: past contributions to atmospheric CO₂ increases and future commitments due to losses of terrestrial sink capacity. *Tellus B* **60**, 583–603 (2008).
- Brovkin, V., Kim, J. H., Hofmann, M. & Schneider, R. A lowering effect of reconstructed Holocene changes in sea surface temperatures on the atmospheric CO₂ concentration. *Glob. Biogeochem. Cycles* **22**, GB1016, doi:10.1029/2006GB002885 (2008).
- Berger, W. H. Increase of carbon dioxide in the atmosphere during deglaciation—the coral-reef hypothesis. *Naturwissenschaften* **69**, 87–88 (1982).
- Ridgwell, A. J., Watson, A. J., Maslin, M. A. & Kaplan, J. O. Implications of coral reef buildup for the controls on atmospheric CO₂ since the Last Glacial Maximum. *Paleoceanography* **18**, 1083–1092 (2003).
- Broecker, W. S., Lynch-Stieglitz, J., Clark, E., Hajdas, I. & Bonani, G. What caused the atmosphere's CO₂ content to rise during the last 8000 years? *Geochim. Geophys. Geosyst.* **2**, 1062, doi:10.1029/2001GC000177 (2001).
- Brovkin, V. *et al.* Carbon cycle, vegetation, and climate dynamics in the Holocene: experiments with the CLIMBER-2 model. *Glob. Biogeochem. Cycles* **16**, doi:10.1029/2001GB001662 (2002).
- Smith, H. J., Fischer, H., Wahlen, M., Mastroianni, D. & Deck, B. Dual modes of the carbon cycle since the Last Glacial Maximum. *Nature* **400**, 248–250 (1999).
- Leuenberger, M., Siegenthaler, U. & Langway, C. C. Carbon isotope composition of atmospheric CO₂ during the last ice-age from an Antarctic ice core. *Nature* **357**, 488–490 (1992).
- Eyer, M. *Highly Resolved $\delta^{13}\text{C}$ Measurements on CO₂ in Air from Antarctic Ice Cores*. 1–113, PhD thesis, Univ. Bern (2004).
- Lourantou, A. *Contraindre l'Augmentation en Dioxyde de Carbone (CO₂) lors des Déglaçiations Basés sur son Rapport Isotopique Stable du Carbone ($\delta^{13}\text{C}$)*. PhD thesis, Univ. Joseph Fourier (2008).
- Loulergue, L. *et al.* New constraints on the gas age-ice age difference along the EPICA ice cores, 0–50 kyr. *Clim. Past* **3**, 527–540 (2007).
- Francey, R. J. *et al.* A 1000-year high precision record of $\delta^{13}\text{C}$ in atmospheric CO₂. *Tellus B* **51**, 170–193 (1999).
- Schurgers, G. *et al.* Dynamics of the terrestrial biosphere, climate and atmospheric CO₂ concentration during interglacials: a comparison between Eemian and Holocene. *Clim. Past* **2**, 205–220 (2006).
- Vecsei, A. & Berger, W. H. Increase of atmospheric CO₂ during deglaciation: constraints on the coral reef hypothesis from patterns of deposition. *Glob. Biogeochem. Cycles* **18**, GB1035, doi:10.1029/2003GB002147 (2004).
- Smith, L. C. *et al.* Siberian peatlands a net carbon sink and global methane source since the early Holocene. *Science* **303**, 353–356 (2004).
- Kim, J. H. *et al.* North Pacific and North Atlantic sea-surface temperature variability during the holocene. *Quat. Sci. Rev.* **23**, 2141–2154 (2004).
- Wang, Y., Mysak, L. A. & Roulet, N. T. Holocene climate and carbon cycle dynamics: experiments with the “green” McGill Paleoclimate Model. *Glob. Biogeochem. Cycles* **19**, GB3022, doi:10.1029/2005GB002484 (2005).
- Beilman, D. W., MacDonald, G. M., Smith, L. C. & Reimer, P. J. Carbon accumulation in peatlands of West Siberia over the last 2000 years. *Glob. Biogeochem. Cycles* **23**, GB1012, doi:10.1029/2007GB003112 (2009).
- Duplessy, J. C. *et al.* Deepwater source variations during the last climatic cycle and their impact on the global deepwater circulation. *Paleoceanography* **3**, 343–360 (1988).
- Crowley, T. J. Ice-age terrestrial carbon changes revisited. *Glob. Biogeochem. Cycles* **9**, 377–389 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is a contribution to the European Project for Ice Coring in Antarctica (EPICA), a joint European Science Foundation/European Commission scientific programme, funded by the EU (EPICA-MIS) and by national contributions from Belgium, Denmark, France, Germany, Italy, the Netherlands, Norway, Sweden, Switzerland and the United Kingdom. The main logistic support was provided by IPEV and PNRA (at Dome C) and AWI (at Dronning Maud Land). We thank A. Landais, D. Rodriguez, E. Capron and G. Dreyfus for the contribution of $\delta^{15}\text{N}$ data as well as P. Nyfeler and K. Grossenbacher for their technical support, T. Tschumi for sharing his Bern3D results, and J. Chappellaz for comments. We acknowledge financial support by the Swiss NSF, the DFG priority programme INTERDYNAMIK and the German climate programme DEKLIM. This is EPICA publication no. 227.

Author Contributions J.E., J.S., D.L., R.S. and M.E. performed the measurements. F.J. performed modelling and interpretation. M.L., H.F. and T.F.S. designed the research. All authors participated in discussions on method development, interpretation and presentation of results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.F.S. (stocker@climate.unibe.ch).

El Niño in a changing climate

Sang-Wook Yeh¹, Jong-Seong Kug¹, Boris Dewitte², Min-Ho Kwon³, Ben P. Kirtman⁴ & Fei-Fei Jin³

El Niño events, characterized by anomalous warming in the eastern equatorial Pacific Ocean, have global climatic teleconnections and are the most dominant feature of cyclic climate variability on subdecadal timescales. Understanding changes in the frequency or characteristics of El Niño events in a changing climate is therefore of broad scientific and socioeconomic interest. Recent studies^{1–5} show that the canonical El Niño has become less frequent and that a different kind of El Niño has become more common during the late twentieth century, in which warm sea surface temperatures (SSTs) in the central Pacific are flanked on the east and west by cooler SSTs. This type of El Niño, termed the central Pacific El Niño (CP-El Niño; also termed the dateline El Niño², El Niño Modoki³ or warm pool El Niño⁵), differs from the canonical eastern Pacific El Niño (EP-El Niño) in both the location of maximum SST anomalies and tropical–midlatitude teleconnections. Here we show changes in the ratio of CP-El Niño to EP-El Niño under projected global warming scenarios from the Coupled Model Intercomparison Project phase 3 multi-model data set⁶. Using calculations based on historical El Niño indices, we find that projections of anthropogenic climate change are associated with an increased frequency of the CP-El Niño compared to the EP-El Niño. When restricted to the six climate models with the best representation of the twentieth-century ratio of CP-El Niño to EP-El Niño, the occurrence ratio of CP-El Niño/EP-El Niño is projected to increase as much as five times under global warming. The change is related to a flattening of the thermocline in the equatorial Pacific.

El Niño statistics exhibits variations on decadal timescales^{7–10}. For instance, the properties of El Niño exhibited frequency and amplitude changes before and after the late 1970s¹⁰. During the late 1990s and 2000s, on the other hand, El Niño events show different characteristics in terms of location of maximum anomalous SST compared to the conventional El Niño^{1–5}. For instance, a prolonged El Niño event during the period of 1990–1994, showed that, in the conventional El Niño region (the far eastern Pacific), the SST anomaly has waxed and waned, while the SST anomaly in the NINO4 region (160°E–150°W, 5°N–5°S) remained positive¹. Other recent studies also argued that there exists a phenomenon in the tropical Pacific that is distinctly different from the canonical El Niño¹¹—this variation¹² of El Niño has a ‘horseshoe’ spatial pattern, flanked by a colder SST on both sides along the Equator^{2–5}. These studies led to various definitions of a new type of El Niño: the dateline El Niño², the El Niño Modoki³, the central Pacific El Niño⁴ and the warm pool El Niño⁵. The El Niño Modoki was named to represent the phenomenon in 2004 that had a maximum SST anomaly in the central tropical Pacific, differing from the conventional El Niño³. In addition, such modification in the structure of El Niño has implications for its teleconnection pattern in many countries surrounding the Pacific Ocean^{2,13,14}. These observations raise the question of whether human-induced global warming¹⁵ can modify our conventional view of El Niño.

We use the historical El Niño indices (the NINO3 SST index and the NINO4 SST index) and the Extended Reconstruction SST data for 1854–2007 to distinguish two variations of El Niño during the boreal winter (December–January–February, DJF). We term these the eastern Pacific El Niño (EP-El Niño) and the central Pacific El Niño (CP-El Niño). These terms have previously been used but with different definitions⁴. Here the terms EP-El Niño and CP-El Niño refer to the years in which the EP-El Niño and the CP-El Niño occurred during winter. Since the 1850s (Supplementary Table 1) the EP-El Niño occurred 32 times and the CP-El Niño occurred 7 times.

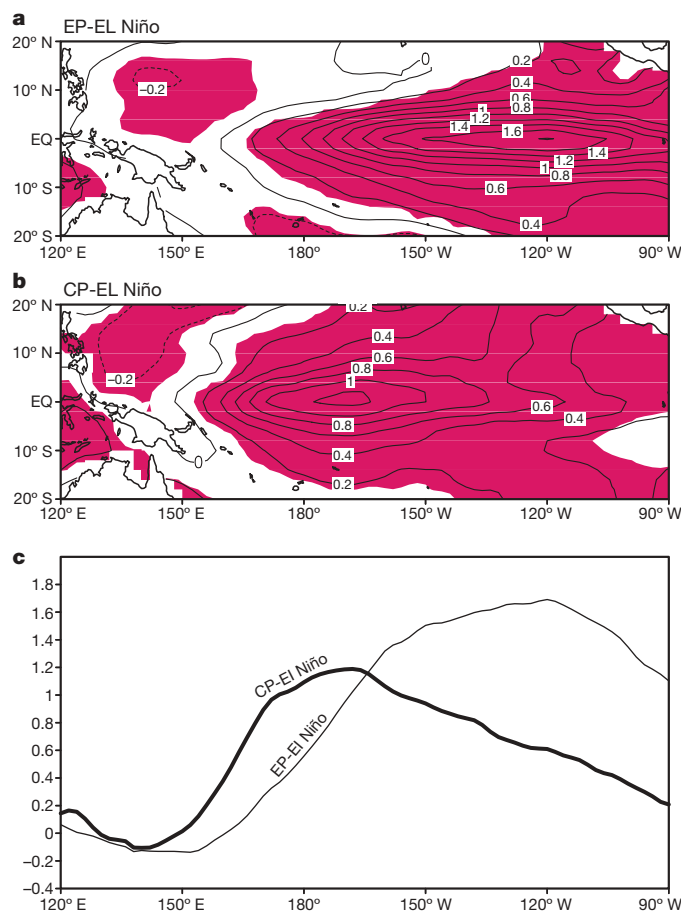


Figure 1 | Deviations of mean SST for the two characteristics of El Niño from the 1854–2006 climatology. **a**, The EP-El Niño; **b**, the CP-El Niño. The contour interval is 0.2 °C and shading denotes a statistical confidence at 95% confidence level based on a Student's *t*-test. **c**, The zonal structure for the composite EP-El Niño (thin line) and CP-El Niño (thick line) averaged over 2°N to 2°S.

¹Climate Change and Coastal Disaster Research Department, Korea Ocean Research and Development Institute, 426-744, Ansan, Korea. ²Laboratoire d'Etude en Géophysique et Oceanographie Spatiale, 14 avenue Edouard Belin, 31400, Toulouse, France. ³Department of Meteorology, School of Ocean and Earth Science and Technology, University of Hawaii, 1680 East-West Road, Honolulu, 96822, Hawaii, USA. ⁴University of Miami, Rosenstiel School of Marine and Atmospheric Science, 4600 Rickenbacker Causeway, Miami, Florida, 33149, USA.

Although the number of CP-El Niño events is relatively small, its frequency increased noticeably after 1990. For the period of 1854–2007, the occurrence ratio of the EP-El Niño before and after 1990 is 0.19 per year and 0.29 per year, respectively, whereas that of the CP-El Niño before and after 1990 is 0.01 per year and 0.29 per year, respectively. Simply put, this result indicates that anomalous warm SSTs in the central equatorial Pacific (that is, CP-El Niño) has been observed more frequently during recent decades³. This result is detectable even if the data are detrended (Supplementary Table 1) and taken from two additional SST data sets (Supplementary Table 3). A profound change in the characteristics of El Niño in recent years is also detectable in an 11-year window sliding correlation coefficients between the two NINO indices (Supplementary Fig. 1).

Figure 1a and b displays the deviation of mean SST for the EP-El Niño and the CP-El Niño from the climatological mean SST (1854–2006). As expected, the EP-El Niño (Fig. 1a) is characterized by maximum anomalous SST in the eastern equatorial Pacific; on the other hand, the centre of maximum SST in the CP-El Niño (Fig. 1b) is located near the dateline in the central equatorial Pacific. The SST composite in Fig. 1b is similar to the previously defined new type of El Niño^{2–5} in spite of an extension of the analysed period and the use of the simple definition of the historical El Niño indices. Figure 1c clearly indicates that the centre of maximum SST of the CP-El Niño is significantly shifted to the west compared to that of the EP-El Niño. The details of the new type of El Niño suggested by previous studies^{1–5} differ slightly from those of the CP-El Niño described here but the overall characteristics are similar.

The large difference of anomalous mean SST between the two types of El Niño results in changes in the total SST pattern in the tropical Pacific (not shown here), which determines the atmospheric response¹⁶. Figure 2a and b displays the composite rainfall corresponding to the EP-El Niño and the CP-El Niño. For the EP-El Niño (Fig. 2a), the centre of maximum anomalous rainfall is observed around the dateline; for the CP-El Niño (Fig. 2b) it is shifted westward to around

165° E. It is clear that anomalous rainfall is largely enhanced in the central and eastern equatorial Pacific and reduced in the western equatorial Pacific during the EP-El Niño compared to the CP-El Niño. Changes in the atmospheric diabatic forcing over the tropics have the potential to modify the tropical–midlatitude teleconnections to the El Niño^{17,18}. Therefore, we would expect the midlatitude response to the EP-El Niño to differ from that of the CP-El Niño, and this has been shown to be true during the last 30 years¹⁴. This is evident from the patterns for anomalous mean atmospheric circulation at 500 hPa in the northern extratropics even over the extended period studied here (Fig. 2c and d) and anomalous mean SST and low-level winds (925 hPa) in the North Pacific (Fig. 2e and f) associated with both types of El Niño. The most striking difference in the teleconnection pattern between the two types of El Niño is in the position of the principal atmospheric centres of action in the extratropics (Fig. 2c and d). In addition, the anomalous North Pacific SST in response to the EP-El Niño and the CP-El Niño is also significantly different (Fig. 2e and f). The spatial manifestation of anomalous SST associated with the EP-El Niño (Fig. 2e) is characterized by cool temperatures in the central North Pacific with an elliptical shape and is accompanied by SST anomalies of the opposite sign to the east, north and south. In contrast to the EP-El Niño, anomalous easterly winds dominate over the central and eastern North Pacific, which may induce anomalous warm SSTs (Fig. 2f). The low-level winds during both types of El Niño are reasonably consistent with the wind–SST interactions in the midlatitudes¹⁹.

Because El Niño and its teleconnections have dramatic societal impacts, such results call for an examination of the El Niño as simulated by the climate models under climate change projections. Here, we examine eleven coupled general circulation models (CGCMs): eleven control runs and eleven climate change runs (Supplementary Table 4). The control run is the twentieth-century climate change model simulation to year 2000 with anthropogenic and natural forcing (that is, 20C3M). The climate change run corresponds to

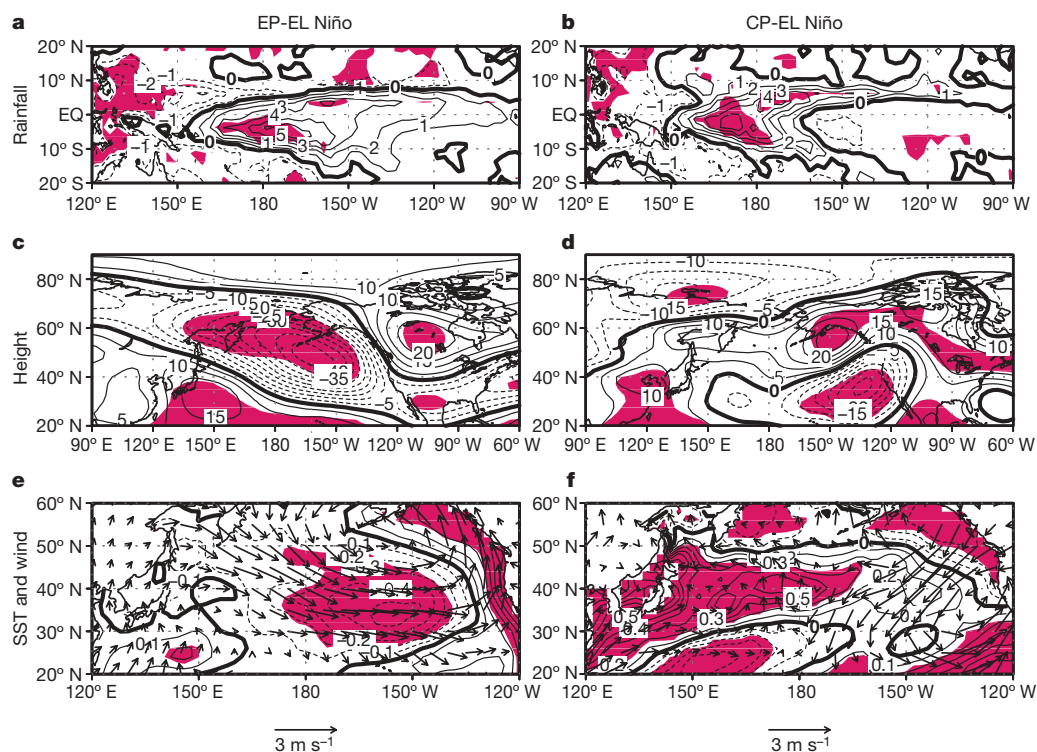


Figure 2 | Deviations for the two characteristics of El Niño from their climatology. **a, b**, The deviation of mean rainfall for the EP-El Niño (**a**) and the CP-El Niño (**b**). The contour interval is 1 mm per day. **c, d**, Mean geopotential height at 500 hPa. The contour interval is 5 m. **e, f**, Mean winds at 925 hPa (arrows, see scale arrow below) and mean SST (line). The solid

(dotted) line denotes positive (negative) deviations from the mean. The contour interval is 0.1 °C. Shading in all panels indicates the region exceeding 95% significance based on a *t*-test and the zero line is denoted by the thick line. The climatology periods are 1979–2006 (for rainfall), 1950–2006 (for geopotential height and winds) and 1854–2006 (for SST), respectively.

the Special Report for Emission Scenario A1B climate change projection (that is, SRESA1B). Here, '20C3M run' refers to data from the 100-years simulation period for the 20C3M run. The term 'SRESA1B run' refers to the last 100 years of the SRESA1B run, in which the concentration of CO₂ is fixed to about 700 p.p.m. We show the ensemble mean composite of the EP-El Niño and the CP-El Niño in the 20C3M run and the SRESA1B run, respectively (Supplementary Figs 2 and 3).

Figure 3 displays the occurrence ratio of the CP-El Niño and EP-El Niño between the control run and the SRESA1B run. Despite the fact that there are discrepancies among CGCMs, it is remarkable that, in eight of 11 models, the occurrence ratio of the CP-El Niño versus the EP-El Niño increases from the 20C3M run to the SRESA1B run. The ensemble mean result for the eleven CGCMs is statistically significant at the 95% confidence level based on the bootstrap method. Furthermore, we test whether the ratio change in each CGCM is significant. The ratio of CP-El Niño to EP-El Niño significantly increases in four of 11 CGCMs at the 95% confidence level, and no other CGCMs show a significant decrease of the occurrence ratio of CP-El Niño to EP-El Niño. Statistical evidence for the increase of CP-El Niño under global warming becomes much stronger when we select the six CGCMs that most realistically capture the occurrence ratio of CP-El Niño to EP-El Niño in the 20C3M run compared to observations (see Supplementary Information). Thus, climate change projections indicate that the CP-El Niño occurs more frequently compared to the EP-El Niño. We also show how the SST variability changes from the 20C3M run to the SRESA1B run in the UKMO-HadCM3 model (Supplementary Fig. 4). We may hypothesize that more frequent CP-El Niño occurrence during recent decades is associated with an anthropogenic climate change. Such changes in El Niño characteristics in future climate are significant enough to modify the tropics–extratropics teleconnection pattern (Supplementary Fig. 5) despite the ability of current models realistically to simulate teleconnections. Furthermore, we expect that such frequent CP-El Niño occurrence under global warming could lead to more effective forcing of drought over India^{3,13,20} and Australia²¹.

Because El Niño dynamics is tightly linked to equatorial ocean mean state²², we argue that such frequent CP-El Niño occurrence is associated with change in the background state under anthropogenic

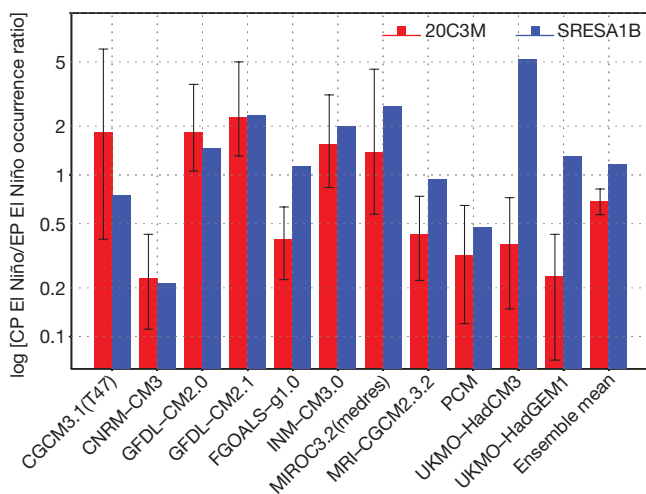


Figure 3 | The CP-El Niño/EP-El Niño occurrence ratio. Red bars, the 20C3M run; blue bars, the SRESA1B run. The vertical error bars denote the upper and lower limits associated with an increase and decrease of the CP-El Niño/EP-El Niño occurrence ratio at the 95% confidence level in the 20C3M run, respectively, based on a bootstrap method. Therefore, there is a significant increase (decrease) of the ratio of the CP-El Niño to the EP-El Niño from the 20C3M run to the SRESA1B run when the blue bar is above (below) the upper (lower) limit of the vertical segment. The y-axis scale is a common logarithmic scale.

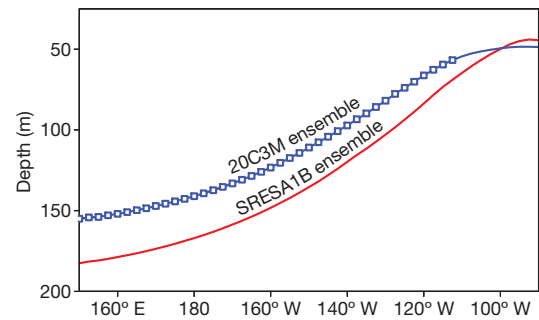


Figure 4 | The ensemble mean thermocline depth. The red line denotes the 20C3M ensemble and the blue line denotes the SRESA1B run in the nine CGCMs: the CGCM3.1(T47), the CNRM-CM3, the GFDL-CM2.0, the GFDL-CM2.1, the FGOALS-g1.0, the MIROC3.2(medres), the MRI-CGCM2.3.2, the PCM and the UKMO-HadCM3. In both runs, the thermocline depth is defined as the depth of the isotherm of the averaged value of temperatures where the vertical gradient of temperature is a maximum along the Equator. The blue data points indicate that the change in the mean thermocline depth from the 20C3M run to the SRESA1B run is significant at the 95% confidence level, based on a *t*-test.

global warming, in particular change in the thermocline structure in the equatorial Pacific. Figure 4 displays the change in mean thermocline depth from the control run to the SRESA1B run. The mean thermocline has risen under global warming in the western-central Pacific, whereas it is slightly deeper in the far eastern Pacific. This results in an overall flattening of the equatorial mean thermocline, which is consistent with a weakened atmospheric Walker circulation and trade winds under global warming²³ and even changes in the thermocline depth during recent decades⁵. In other words, the SST warms as a result of thermal forcing, which leads to weaker easterlies and enhanced poleward Sverdrup transport and hence a shoaling of the thermocline depth. How might this affect the stability of the CP-El Niño?

We can understand this destabilizing process in terms of the two important feedback processes associated with El Niño dynamics, that is, thermocline feedback versus zonal advective feedback. Although the trade winds reduce under global warming, this reduces upwelling and thus the thermocline feedback. In contrast, a shallower thermocline in the central Pacific, as in the SRESA1B run, tends to enhance the SST anomaly induced by vertical advection there (because isotherm vertical displacements within the thermocline depth can more easily influence the SST). In addition, such a shallowing thermocline tends to dominate the zonal advective feedback in the central Pacific, which may promote a more intense CP-El Niño^{5,22,24}. Overall, the change in thermocline structure from the 20C3M run to the SRESA1B run is consistent with the increased variability of the SST anomaly in the central Pacific. This physical consistency fits with the result reported here: the probable increased occurrence of the CP-El Niño under global warming.

METHODS SUMMARY

The two kinds of El Niño were diagnosed from observations and eleven models of the Program for Climate Model Diagnosis and Intercomparison (PCMDI). We propose a classification based on the historical NINO3 and NINO4 SST indices during winter and inferred from composite analyses to distinguish the CP-El Niño from the EP-El Niño. Applied to the simulation for the present (20C3M) and for the future (SRESA1B), we derived a projection of the occurrence ratio of CP-El Niño to EP-El Niño. See the Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 December 2008; accepted 21 July 2009.

- Latif, M., Kleeman, R. & Eckert, C. Greenhouse warming, decadal variability, or El Niño? An attempt to understand the anomalous 1990s. *J. Clim.* **10**, 2221–2239 (1997).

2. Larkin, N. K. & Harrison, D. E. Global seasonal temperature and precipitation anomalies during El Niño autumn and winter. *Geophys. Res. Lett.* **32**, L13705, doi:10.1029/2005GL022738 (2005).
3. Ashok, K., Behera, S. K., Rao, S. A., Weng, H. & Yamagata, T. El Niño Modoki and its possible teleconnection. *J. Geophys. Res.* **112**, C11007, doi:10.1029/2006JC003798 (2007).
4. Kao, H.-Y. & Yu, J.-Y. Contrasting Eastern-Pacific and Central-Pacific types of ENSO. *J. Clim.* **22**, 615–632 (2009).
5. Kug, J.-S., Jin, F.-F. & An, S.-I. Two types of El Niño events: cold tongue El Niño and warm pool El Niño. *J. Clim.* **22**, 1499–1515 (2009).
6. Meehl, G. A. *et al.* The WCRP CMIP3 multimodel dataset. *Bull. Am. Meteorol. Soc.* **88**, 1383–1394 (2007).
7. Trenberth, K. & Shea, D. J. On the evolution of the Southern Oscillation. *Mon. Weath. Rev.* **115**, 3078–3096 (1987).
8. Cobb, K., Charles, C., Cheng, H. & Edwards, R. El Niño/Southern Oscillation and tropical Pacific climate during the last millennium. *Nature* **424**, 272–276 (2003).
9. An, S.-I. & Jin, F.-F. Nonlinearity and asymmetry of ENSO. *J. Clim.* **17**, 2399–2412 (2004).
10. An, S.-I. & Wang, B. Interdecadal change of the structure of the ENSO mode and its impact on the ENSO frequency. *J. Clim.* **13**, 2044–2055 (2000).
11. Rasmusson, E. M. & Carpenter, T. H. Variations in tropical sea surface temperature and surface wind fields associated with the southern oscillation/El Niño. *Mon. Weath. Rev.* **110**, 354–384 (1982).
12. Trenberth, K. E. & Stepaniak, D. P. Indices of El Niño evolution. *J. Clim.* **14**, 1697–1701 (2001).
13. Weng, H., Ashok, K., Behera, S. K., Rao, S. A. & Yamagata, T. Impacts of recent El Niño Modoki on dry/wet conditions in the Pacific Rim during boreal summer. *Clim. Dyn.* **29**, 113–129 (2007).
14. Weng, H., Behera, S. K. & Yamagata, T. Anomalous winter climate conditions in the Pacific Rim during recent El Niño Modoki and El Niño events. *Clim. Dyn.* **32**, 663–674 (2009).
15. Solomon, S. *et al.* (eds) *Climate Change 2007: The Physical Science Basis* (Cambridge University Press for the Intergovernmental Panel on Climate Change, 2007).
16. Hoerling, M. P. & Kumar, A. Atmospheric response patterns associated with tropical forcing. *J. Clim.* **15**, 2184–2203 (2002).
17. Alexander, M. *et al.* The atmospheric bridge: The influence of ENSO teleconnections on air-sea interaction over the global oceans. *J. Clim.* **15**, 2205–2228 (2002).
18. Barsugli, J. & Sardeshmukh, P. D. Global atmospheric sensitivity to tropical SST anomalies throughout the Indo-Pacific basin. *J. Clim.* **15**, 3427–3442 (2002).
19. Cayan, D. R. Latent and sensible heat flux anomalies over the northern oceans: driving the sea surface temperature. *J. Phys. Oceanogr.* **22**, 859–881 (1992).
20. Kumar, K. K., Rajagopalan, B., Hoerling, M., Bates, G. & Cane, M. Unraveling the mystery of Indian monsoon failure during El Niño events. *Science* **314**, 115–119 (2006).
21. Wang, G. & Hendon, H. H. Sensitivity of Australian rainfall to inter-El Niño variations. *J. Clim.* **20**, 4211–4226 (2007).
22. Fedorov, A. V. & Philander, S. G. H. Is El Niño changing? *Science* **288**, 1997–2002 (2000).
23. Vecchi, G. A. *et al.* Weakening of tropical Pacific atmospheric circulation due to anthropogenic forcing. *Nature* **441**, 73–76 (2006).
24. Bejarano, L. & Jin, F.-F. Coexistence of equatorial coupled mode of ENSO. *J. Clim.* **21**, 3051–3067 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the international modelling groups for providing their data and PCMDI and the IPCC Data Archive at LLNL/DOE for collecting, archiving and making the data readily available. S.-W.Y. and J.-S.K. are supported by KORDI (grants PE98401, PP00720). B.D. benefited from funding from the PCCC project (Peru Chile Climate Change) of the ANR (Agence Nationale de la Recherche). J.-S. K. and F.-F. J. are also supported by NSF grants ATM 060552 and AMT 065145 and NOAA grant GC01-229.

Author Contributions S.-W.Y., M.K. and J.-S.K. contributed to analysis. S.-W.Y., J.-S.K., B.D., B.P.K. and F.-F.J. contributed to writing the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.-W.Y. (swyeh@kordi.re.kr).

METHODS

The SSTs analysed in this study are taken from the Extended Reconstruction SST version 2 (ERSST.v2) covering the period of 1854–2007 released by the National Climatic Data Center²⁵. In addition, the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) data are used for the period of 1979–2007 (ref. 26). Atmospheric circulation data were taken from the National Centers for Environmental Prediction/National Center for Atmospheric Science (NCEP/NCAR) reanalysis data²⁷ which use a grid with a horizontal resolution of $2.5^\circ \times 2.5^\circ$. To define the two types of El Niño we first collect the years in which the NINO3 SST index during the boreal winter (DJF) is above 0.5°C or the NINO4 SST index during winter is above 0.5°C . The DJF NINO3 SST index is defined by the time series of the seasonal (that is, DJF) mean SST anomaly averaged over the NINO3 region (150°W – 90°W , 5°N – 5°S). Similarly, the DJF NINO4 SST index is the same as the DJF NINO3 SST index except for the NINO4 region (160°E – 150°W , 5°N – 5°S). The seasonal mean SST anomaly is defined as seasonal mean deviations from a climatological (1854–2006) seasonal mean SST. Of those years, an EP-El Niño year is defined as a year in which the DJF NINO3 SST index is greater than the DJF NINO4 SST index. On the other hand, a CP-El Niño year is defined as a year in which the DJF NINO4 SST index is greater than the DJF NINO3 SST index. Using this classification, the composite for mean precipitation, 500 hPa geopotential height and surface winds is derived for the two types of El Niño. The seasonal mean anomalies for these variables are also defined as seasonal mean deviations from a climatological seasonal mean.

The method is further applied to eleven CGCM simulations in the 20C3M run and the SRESA1B run made by the Program for Climate Model Diagnosis and Intercomparison (PCMDI). The occurrence ratio of CP-El Niño to EP-El Niño is derived and the change in statistics of this parameter from the 20C3M run to the

SRESA1B run is examined. To examine whether the change in the CP-El Niño/EP-El Niño occurrence ratio from the SRESA1B run is significantly different from the internal variability of the 20C3M run, we constructed the probability distribution function of the internal variability for the occurrence ratio from each individual model in the 20C3M run using a bootstrap method²⁸. First, we randomly select N El Niño events of the total El Niño events for each model in the 20C3M run. During the random selection process, overlapping selection is allowed, so that one El Niño event can be selected again. Note that N is a total number of the CP-El Niño and EP-El Niño events simulated from an individual model; therefore, N differs for each model. From the selected N events in each model, we separate them into the CP-El Niño and EP-El Niño events and then we calculate CP-El Niño/EP-El Niño occurrence ratio. By repeating this process 10,000 times, we obtain 10,000 values for the occurrence ratio and the probability distribution function for the occurrence ratio is constructed for each model. The top and bottom of each error bar in Fig. 3 represents the 2.5% and 97.5% ranking from the probability distribution function, respectively, indicating the 95% confidence level. If the occurrence ratio from the SRESA1B run is out of the range of the 2.5% and 97.5% ranking, it indicates that the change in occurrence ratio from the 20C3M run to the SRESA1B run is significant at the 95% confidence level.

25. Smith, T. M. & Reynolds, R. W. Improved extended reconstruction of SST (1854–1997). *J. Clim.* **17**, 2466–2477 (2004).
26. Xie, P., & Arkin, P. A. Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical outputs. *Bull. Am. Meteorol. Soc.* **78**, 2539–2558 (1997).
27. Kalnay, E. *et al.* The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–471 (1996).
28. Efron, B. *The Jackknife, the Bootstrap, and Other Resampling Plans* 1–92 (Society for Industrial and Applied Mathematics, 1982).

An epistatic ratchet constrains the direction of glucocorticoid receptor evolution

Jamie T. Bridgman¹, Eric A. Ortlund³ & Joseph W. Thornton^{1,2}

The extent to which evolution is reversible has long fascinated biologists^{1–8}. Most previous work on the reversibility of morphological and life-history evolution^{9–13} has been indecisive, because of uncertainty and bias in the methods used to infer ancestral states for such characters^{14,15}. Further, despite theoretical work on the factors that could contribute to irreversibility^{1,8,16}, there is little empirical evidence on its causes, because sufficient understanding of the mechanistic basis for the evolution of new or ancestral phenotypes is seldom available^{3,8,17}. By studying the reversibility of evolutionary changes in protein structure and function, these limitations can be overcome. Here we show, using the evolution of hormone specificity in the vertebrate glucocorticoid receptor as a case-study, that the evolutionary path by which this protein acquired its new function soon became inaccessible to reverse exploration. Using ancestral gene reconstruction, protein engineering and X-ray crystallography, we demonstrate that five subsequent ‘restrictive’ mutations, which optimized the new specificity of the glucocorticoid receptor, also destabilized elements of the protein structure that were required to support the ancestral conformation. Unless these ratchet-like epistatic substitutions are restored to their ancestral states, reversing the key function-switching mutations yields a non-functional protein. Reversing the restrictive substitutions first, however, does nothing to enhance the ancestral function. Our findings indicate that even if selection for the ancestral function were imposed, direct reversal would be extremely unlikely, suggesting an important role for historical contingency in protein evolution.

Evolutionary reversibility represents a strong test of the importance of contingency and determinism in evolution. If selection is limited in its ability to drive the reacquisition of ancestral forms, then the future outcomes available to evolution at any point in time must depend strongly on the present state and, in turn, on the past^{2,4,8,18}. Ready reversibility, in contrast, would indicate that natural selection can produce the same optimal form in any given environment, irrespective of history¹⁹. The evolutionary reversibility of a protein can be evaluated at three levels: molecular sequence, protein function, and the structural/mechanistic underpinnings for that function. The latter is most relevant to understanding the roles of contingency and determinism in evolution. Exact molecular reversal to the ancestral amino acid sequence is extremely unlikely and of trivial interest, because of the large number of sequences that code for the same structure and function. Selection will always produce adaptive functions or phenotypes in some form; however, if the underlying mechanism for a reversed function differs from that of the ancestor, then a new, analogous state will have been achieved by onward evolution, not reversal—a situation similar to false morphological reversal caused by convergent evolution⁴. True reversal, involving restoration of the ancestral phenotype by the ancestral structure–function relations, would indicate that the

forms of functional proteins can evolve deterministically, irrespective of contingent historical events.

Recent developments in techniques for studying protein evolution allow the reversibility of protein structure and function to be studied directly. The intrinsic functions and atomic structures of ancestral genes can be determined by inferring their sequences using maximum-likelihood phylogenetics, then biochemically synthesizing, expressing and characterizing them using functional assays and X-ray crystallography²⁰. The mechanisms by which new functions evolved can be identified by introducing historical substitutions into ancestral backgrounds and characterizing their effects on structure and function^{21,22}. Using these techniques, we recently established the mechanistic basis for the evolution of a new function in the glucocorticoid receptor, a DNA-binding transcription factor activated by the steroid hormone cortisol to regulate the long-term stress response and other processes in humans and other vertebrates^{23,24}. Specifically, we showed that the cortisol-specificity of the glucocorticoid receptor ligand-binding domain (LBD) evolved from a more promiscuous ancient receptor that was activated by the mineralocorticoids aldosterone and deoxycorticosterone (DOC) and, albeit more weakly, by cortisol. The new specificity of glucocorticoid receptors (Fig. 1a–c) evolved because of a marked change in structure–function relations during the 40-million-year period between AncGR1 (the glucocorticoid receptor in the last common ancestor of cartilaginous and bony fish, which had the ancestral function) and AncGR2 (the glucocorticoid receptor in the ancestor of tetrapods and ray-finned fish, which was cortisol-specific). Of the 37 amino acid changes that occurred during this interval, two conserved substitutions (Ser106Pro and Leu111Gln, called group X for convenience) were necessary and sufficient to switch the preference of the resurrected AncGR1 from mineralocorticoids to cortisol. Ser106Pro radically repositioned helix 7 along the edge of the ligand pocket, reducing activation by all hormones but moving site 111 closer to the ligand. In this new position, the hydrophobic substitution Leu111Gln generated a new hydrogen bond to the 17-hydroxyl group unique to cortisol, specifically restoring sensitivity to that hormone (Fig. 1d). Three more conserved substitutions (group Y) completed the loss of mineralocorticoid sensitivity to yield a cortisol-specific receptor; these changes further destabilized the receptor complex with mineralocorticoids but enhanced interaction with the 17-hydroxyl of cortisol. The protein could not tolerate group Y, however, without an additional pair of permissive substitutions (group Z), which added stability to the structural elements destabilized by group Y and conferred the full glucocorticoid-receptor-like function upon AncGR1 + XYZ (Fig. 1e and ref. 24).

The most direct pathway to reverse the evolution of the structure and function of AncGR2 would be to reverse the key substitutions that generated the derived phenotype during the ‘forward’ evolution of AncGR1. We used site-directed mutagenesis on AncGR2 to reverse

¹Center for Ecology and Evolutionary Biology, and ²Howard Hughes Medical Institute, University of Oregon, Eugene, Oregon 97403, USA. ³Biochemistry Department, Emory University School of Medicine, Atlanta, Georgia 30322, USA.

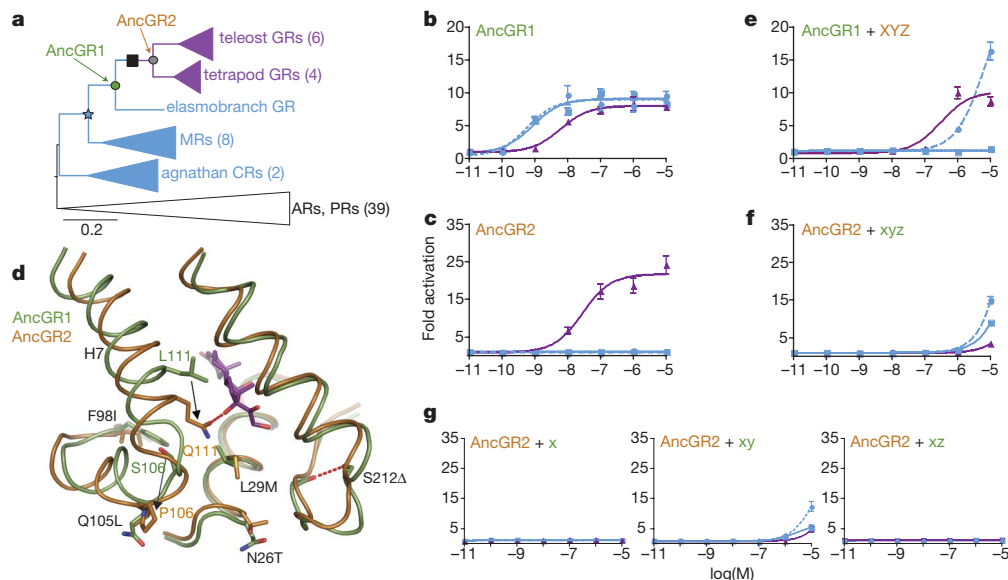


Figure 1 | Evolution and reversibility of glucocorticoid receptor function.

a, Reduced phylogeny of corticosteroid receptors. Blue, receptors sensitive to both cortisol and mineralocorticoids; purple, sensitive to cortisol only; black, other steroid receptors (AR, androgen receptor; PR, progesterone receptor). CR, cortisol receptor; GR, glucocorticoid receptor; MR, mineralocorticoid receptor. Ancestral proteins AncGR1 and AncGR2 are labelled. Thirty-seven amino acid changes, including groups X, Y and Z, occurred during the interval between these two proteins (black box; for complete list and alignment, see Supplementary Fig. 1). Parentheses show the number of sequences in each group. Scale bar for branch lengths is shown in substitutions per site. **b, c**, Ligand sensitivities of AncGR1 (**b**) and AncGR2 (**c**), shown as the fold increase in expression of a luciferase reporter in the presence of increasing doses of cortisol (purple), aldosterone (solid blue),

the amino acids in groups X, Y and Z to their ancestral states (x, y and z), and characterize their effect on receptor function in a luciferase reporter gene assay (Fig. 1f). Surprisingly, AncGR2-xyz was unable to activate transcription in response to any ligand, even at very high concentrations. Reversing only the large-effect mutations in group X also produced a non-functional AncGR2-x receptor, as did all combinations that included restoration of group x (Fig. 1g). These results indicate that the structure and function of AncGR2 are not reversible through this direct path, because the ancestral amino acids in group x—and the change in conformation they cause—are incompatible with the derived background. These states, however, were present in AncGR1 just 40 million years earlier and have been conserved in all mineralocorticoid receptors ever since, indicating that further epistatic modifiers must have evolved between AncGR1 and AncGR2.

To identify candidate historical substitutions for this restrictive effect, we combined phylogenetic and structural analysis. Thirty substitutions in addition to X, Y and Z occurred during the AncGR1–AncGR2 interval. We reasoned that amino acids required for the ancestral function should be conserved in the AncGR1-like state in extant receptors that retain the ancestral sensitivity to DOC and aldosterone; unlike X, Y and Z, however, they would not be expected to be conserved in the glucocorticoid-receptor-like receptors. Of the 30 remaining sites that changed between AncGR1 and AncGR2, only six were invariant among all or all but one of the DOC/aldosterone-sensitive receptors (Fig. 2a and Supplementary Fig. 1). To predict which of these were most likely to enable the ancestral function, we expressed the AncGR2 LBD and used X-ray crystallography (Supplementary Table 1) to determine its empirical atomic structure at 2.5 Å resolution in complex with the synthetic glucocorticoid dexamethasone (Fig. 2b and Supplementary Figs 2 and 3). The monomeric AncGR2 structure adopts the canonical active conformation for nuclear receptors²⁵ and is nearly identical to the AncGR2 structure previously predicted by homology modelling²⁴.

and DOC (dashed blue). **d**, Conformational change causing cortisol-specificity in AncGR2 (see ref. 24). Partial structures of AncGR1 and AncGR2 are superimposed. Substitutions in group X (Ser106Pro and Leu111Gln) are large effect mutations that reposition helix 7 (H7) and form a hydrogen bond to the 17-OH that is unique to cortisol (purple). Black arrows indicate change in position of these residues. Substitutions in groups Y (Leu29Met, Phe98Ile and Ser212A), and Z (Asn26Thr and Gln105Leu) optimize the derived function. **e**, When substitutions in sets X, Y and Z are introduced into AncGR1, they recapitulate the evolution of a cortisol-specific activator. **f**, When these substitutions are reversed to the ancestral state (xyz) in the AncGR2 background, activation by all ligands is lost. **g**, All AncGR2 combinations in which group X is reversed also yield non-functional receptors. All error bars denote s.e.m.

Five of the six candidate substitutions identified by phylogenetic analysis (group W: His84Gln, Tyr91Cys, Ala107Tyr, Gly114Gln and Leu197Met) are on or interact directly with the repositioned helix 7; the other (Val234Phe) is far from the remodelled portion of the protein and does not seem to interact with it directly or indirectly (Fig. 2b). By comparing the structure of AncGR2 with that of AncGR1 (ref. 24), we predicted that the derived states at these five sites would be incompatible with the ancestral structure and function, because the AncGR1 states stabilize the active conformation with helix 7 in its ancestral position (Fig. 2b, c), but those in AncGR2 fail to support this conformation or actively clash with it. Specifically (Fig. 2c), two of these residues in AncGR1 (Gly 114 and Leu 197) are close together and allow tight packing of helix 7 in the ancestral position against helix 10. AncGR2, in contrast, contains Gln and Met at these positions, the side chains of which are longer and less hydrophobic; the repositioning of helix 7 allows these two residues to be tolerated, but in the ancestral conformation their side chains would clash, pushing the two helices apart and away from the ligand. A second pair, the aromatic residues His 84 and Tyr 91, form a pi-stack in AncGR1, stabilizing the β -strand that abuts the ligand and helix 7. Substitution of these residues to Gln and Cys (as in AncGR2) would destroy this interaction, increasing flexibility in the ligand pocket and destabilizing the active complex; AncGR2 can presumably tolerate this effect because of the extra stability contributed by the hydrogen bond between Gln 111 and cortisol. The fifth candidate site, Ala 107, lies near the base of helix 7, at the mouth of the ligand pocket where helices 3, 7 and 10 pack together. Replacement of Ala 107 with the bulky tyrosine of AncGR2 would clash with helices 3 and 10; however, the movement of helix 7 in AncGR2 repositions site 107, allowing a tyrosine to be tolerated.

To test the hypothesis that group W substitutions impede direct reversal of the key function-switching mutations, we used site-directed mutagenesis to reverse group W in the AncGR2-xyz background. We

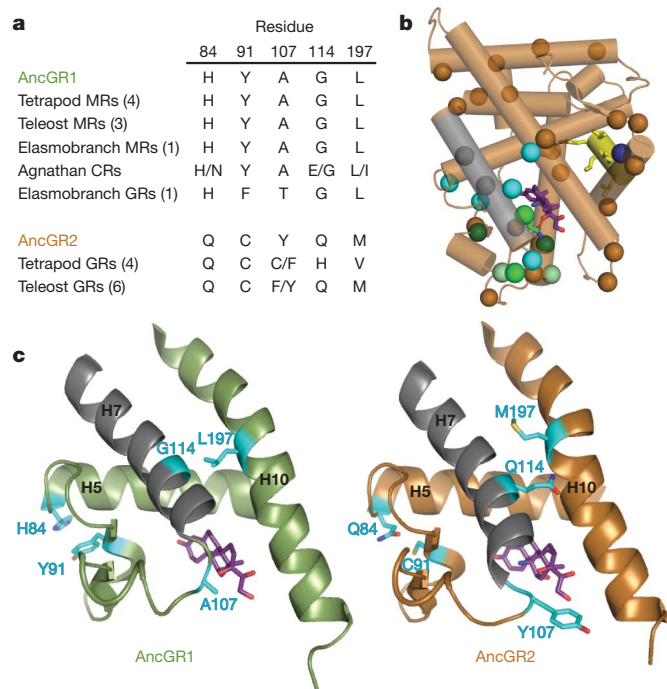


Figure 2 | Identification of restrictive substitutions that impede reversibility. **a**, Group W residues are conserved in the AncGR1-like state in virtually all extant receptors that retain the ancestral function. Parentheses show the number of sequences in each group. **b**, X-ray crystal structure (PDB accession 3GN8) of AncGR2 (bronze) with dexamethasone (purple). Repositioned helix 7 is shown in grey. Residues substituted between AncGR1 and AncGR2 are marked with spheres at the α -carbon. Cyan, candidate restrictive substitutions (group W). Sites in groups X, Y and Z are shown in medium, dark and light green, respectively. Blue, Val234Phe. **c**, Ligand pockets of AncGR1 (green, with cortisol) and AncGR2 (bronze, with dexamethasone). When in their ancestral states, group w residues (cyan) are predicted to support the ancestral conformation of helix 7 (grey), but to destabilize that conformation when in the derived states found in AncGR2.

then determined their functional effects using a luciferase reporter assay (Fig. 3). As predicted, reversing all five group W mutations restored the ancestral phenotype, yielding a sensitive, promiscuous receptor with a nanomolar response to both mineralocorticoids and cortisol and, like AncGR1, a preference for aldosterone (Fig. 3a). All five group W substitutions contribute to AncGR2's intolerance of the ancestral structure/function: Tyr107Ala alone partially rescued the transcriptional function of AncGR2-xyz and shifted it substantially towards the ancestral promiscuous phenotype, as did the pairs Gln84His/Cys91Tyr and Gln114Gly/Met197Leu (Fig. 3b). Restoring the single mutations Gln84His, Cys91Tyr, Gln114Gly and Met197Leu had no or very weak effects (Supplementary Fig. 4), presumably because of the structural interactions within each pair required to improve the receptor's function.

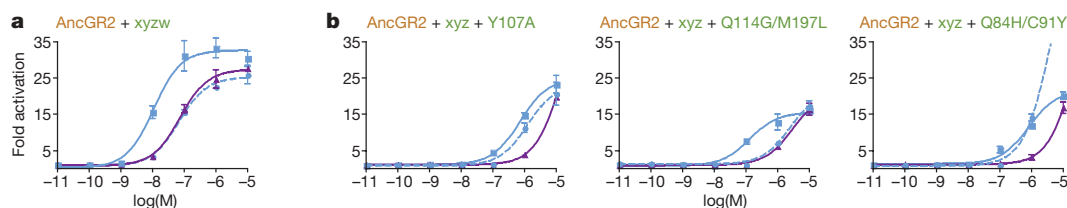


Figure 3 | Restrictive substitutions impede evolutionary reversibility. **a**, When group W substitutions are restored to their ancestral state (w), the non-functional AncGR2-xyz is rescued, and the ancestral sensitivity to all three ligands is restored. Fold increase in luciferase reporter expression is shown with cortisol (purple), aldosterone (solid blue), and DOC (dashed

blue). **b**, Group W substitutions all impede reversibility: restoring the ancestral states singly (Tyr107Ala) or in structurally interacting pairs (Gln84His/Cys91Tyr and Gln114Gly/Met197Leu) partially rescues AncGR2-xyz. Error bars, s.e.m.

To test the hypothesis that group W substitutions specifically undermine the stability of the ancestral helix 7 conformation, we restored the ancestral state (w) in all possible combinations of x, y and z in the AncGR2 background (Fig. 4 and Supplementary Fig. 5). As predicted, reversal to x always impairs both the ancestral and derived functions unless group w has been reversed first. Taken together, our experiments indicate that these five mutations prevent direct evolutionary reversal by weakening aspects of the receptor structure that were required to support the ancestral conformation. By reversing all of these restrictive substitutions, the ancestral structure and function can be largely restored. The reversed AncGR2-xyzw remains slightly less sensitive to hormones than AncGR1, indicating that some of the other 25 substitutions during the AncGR1–AncGR2 interval make further, minor contributions to impeding direct evolutionary reversal (Fig. 4a, b). The restrictive effect of mutations in group W on the reversal of group X does not depend on whether these other 25 substitutions are in their ancestral or derived states (Fig. 4a, b and Supplementary Fig. 5). Although there are other combinations of individual substitutions that we did not test, our results indicate that neither the restrictive effect of group W mutations nor the permissive effect of reversing them depends narrowly on a specific genetic background.

The restrictive mutations that impede direct reversal may have been adaptive or neutral when they occurred. To characterize the ‘forward’ effect of group W mutations on receptor function, we recapitulated them in the AncGR1 background with various combinations of groups X, Y and Z (Fig. 4b). In AncGR1-XYZ and other X-containing backgrounds, W mutations cause a weak or moderate improvement in receptor activation and cortisol-specificity, presumably by stabilizing the derived position of helix 7 and its interaction with cortisol. In the AncGR1 background and all other combinations that include x, however, W mutations markedly reduce sensitivity to all hormones. Because selection makes evolutionary trajectories that pass through non-functional intermediates far less likely than those involving functional intermediates at every step²⁶, W mutations are unlikely to have been complete before the remodelling and functional shift triggered by group X. Once X was in place, however, the W mutations that prevent direct evolutionary reversal probably optimized the derived function or were neutral.

Our findings indicate that epistatic modifiers—at least some of which occurred after the new function of the glucocorticoid receptor evolved—acted as an evolutionary ratchet, making re-evolution of the ancestral structure–function relations far more difficult than it was initially. Reversal by a direct path that restores the key residues in group X became exceedingly unlikely, because features that once enabled the ancestral conformation of helix 7 had been modified. To restore the ancestral conformation by reversing group X, the restrictive effect of the substitutions in group W must first be reversed, as must group Y (Fig. 4a, b). Reversal to w and y in the absence of x, however, does nothing to enhance the ancestral function; in most contexts, reversing these mutations substantially impairs both the ancestral and derived functions (Fig. 4a, b). Furthermore, the permissive effect of reversing four of the mutations

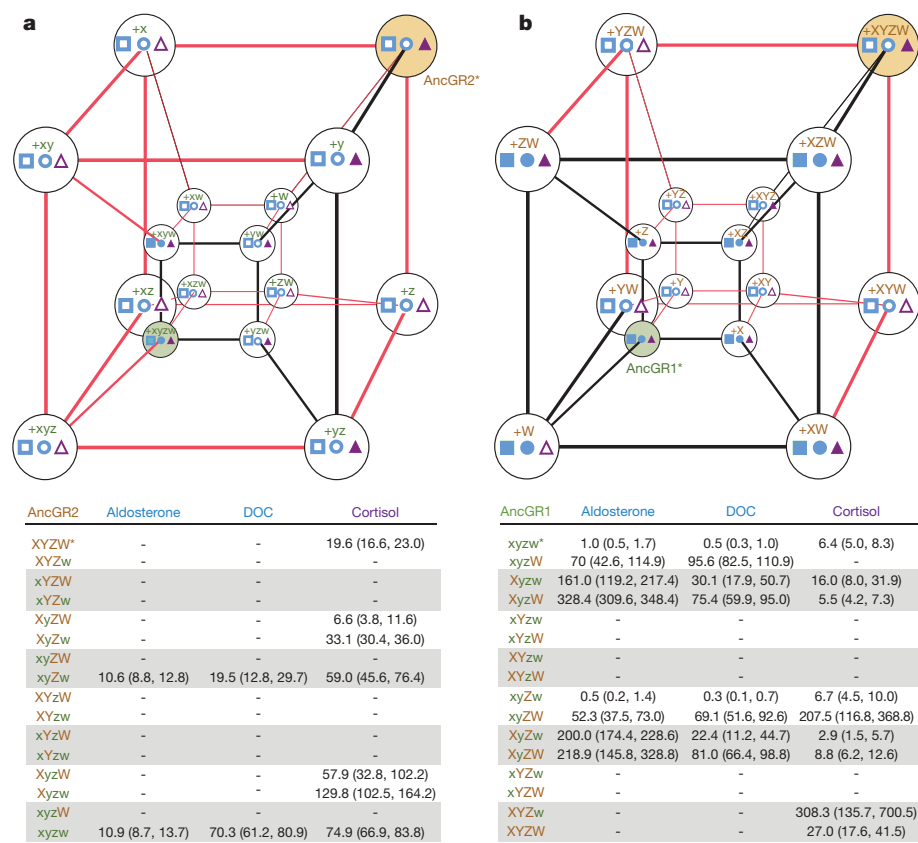


Figure 4 | Epistasis limits trajectories of reverse and forward evolution. The corners of each hypercube represent states for residue sets X, Y, Z and W. Edges show pathways between the derived (XYZW, bronze) and ancestral (xyzw, green) states. Red edges show unlikely evolutionary paths through non-functional intermediates; black paths pass through functional intermediates. Filled shapes at vertices indicate sensitivity to aldosterone (blue squares), DOC (blue circles), and cortisol (purple triangles); empty shapes, no activation by these hormones. Tables below each cube show sensitivity to each hormone as the EC₅₀ (concentration required for

half-maximal activation), with 95% confidence interval shown in parentheses. Dashes, no activation; asterisks, state combinations in AncGR2 and AncGR1. **a**, Limited evolutionary pathways to reverse AncGR2 (bronze) to the ancestral structure and function. Mutations were introduced in the AncGR2 background. **b**, Functional effect of substitutions during ‘forward’ evolution when introduced into AncGR1 (green). The sets X, Y, Z and W each contain more than one site, so the complete sequence space for each cube has 12 dimensions and 4,096 vertices.

in group W requires pairs of substitutions at interacting sites. Selection for the ancestral function would therefore not be sufficient to drive AncGR2 back to the ancestral states of w and x, because passage through deleterious and/or neutral intermediates would be required; the probability of each required substitution would be low, and the probability of all in combination would be virtually zero. We have examined the sufficiency of selection to drive direct evolutionary reversal. There may be other potentially permissive mutations, of unknown number, that could compensate for the restrictive effect of group W and allow the ancestral conformation to be restored. Reversal by such indirect pathways could be driven by selection, however, only if these other mutations, unlike those we studied, could somehow relieve the steric clashes and restore the lost stabilizing interactions that make the ancestral position of helix 7 intolerable in AncGR2, and also independently restore the ancestral function when helix 7 is in its radically different derived conformation. Whether or not mutations that could achieve these dual ends exist, reversal to the ancestral conformation would require a considerably more complex pathway than was necessary before the ratchet effect of W evolved. The extent to which our observations concerning the evolutionary reversibility of glucocorticoid receptors can be generalized to other proteins requires further research. We predict that future investigations, like ours, will support a molecular version of Dollo’s law⁴: as evolution proceeds, shifts in protein structure–function relations become increasingly difficult to reverse whenever those shifts have complex architectures, such as requiring conformational changes or

epistatically interacting substitutions^{2,8,16}. Phenotypes at higher levels of genetic organization may also display ratchet-like modes of evolution if optimization of a derived phenotype involves changes in one gene, regulatory element, morphological structure, or developmental process that epistatically undermine the conditions that enabled the ancestral state at other such ‘loci’. In contrast, phenotypic shifts caused by single or additive genetic changes are likely to be readily reversible^{1,27}. Our observations suggest that history and contingency during glucocorticoid receptor evolution strongly limited the pathways that could be deterministically followed under selection. The ‘adaptive peak’ represented by the promiscuous AncGR1 is a relatively close neighbour in sequence space to the more specific AncGR2. This peak was occupied in the ancestor of jawed vertebrates—indicating that no intrinsic constraints prevent its realization—but it became far more difficult to access just 40 million years later because of intervening epistatic mutations. Selection is an extraordinarily powerful evolutionary force;²⁸ nevertheless, our observations suggest that, because of the complexity of glucocorticoid receptor architecture, low-probability permissive substitutions were required to open some mutational trajectories to exploration under selection^{24,29}, whereas restrictive substitutions closed other potential paths. Under selection, some kind of adaptation will always occur³⁰, but the specific adaptive forms that are realized depend on the historical trajectory that precedes them. The conditions that once facilitated evolution of the glucocorticoid receptor’s ancestors were destroyed during the realization of its present form^{2,4,7,16,18}. The past is difficult to recover because it was built on the foundation of its

own history, one irrevocably different from that of the present and its many possible futures.

METHODS SUMMARY

Peptide sequences of the AncGR1 and AncGR2 LBDs were inferred using maximum-likelihood phylogenetics from an alignment of 60 peptide sequences of extant steroid and related receptors as previously described²⁴. Complementary DNAs coding for these peptides were synthesized and subcloned and expressed as fusion constructs with Gal4-DBD (DNA-binding domain) in Chinese hamster ovary (CHO-K1) cells. Activation was measured using dual luciferase assays in the presence of increasing concentrations of various hormones. AncGR2-LBD was bacterially expressed as a maltose-binding protein/TEV fusion in the presence of dexamethasone, then purified, cleaved, dialysed, concentrated, crystallized and diffracted using X-ray crystallography at the Advanced Photon Source. The atomic structure was determined to 2.5 Å by molecular replacement based on the previously described human glucocorticoid-receptor-based homology model of AncGR2-LBD²⁴, followed by further refinement. Details are presented in Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 May; accepted 30 June 2009.

- Muller, H. J. Reversibility in evolution considered from the standpoint of genetics. *Biol. Rev. Camb. Philos. Soc.* **14**, 261–280 (1939).
- Simpson, G. G. *The Major Features of Evolution* 310–312 (Columbia Univ. Press, 1953).
- Crick, F. H. The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379 (1968).
- Gould, S. J. Dollo on Dollo's law: irreversibility and the status of evolutionary laws. *J. Hist. Biol.* **3**, 189–212 (1970).
- Dobzhansky, T. G. *Genetics of the Evolutionary Process* 428–430 (Columbia Univ. Press, 1971).
- Macbeth, N. Reflections on irreversibility. *Syst. Zool.* **29**, 402–404 (1980).
- Bull, J. J. & Charnov, E. L. On irreversible evolution. *Evolution* **39**, 1149–1155 (1985).
- Teotonio, H. & Rose, M. R. Perspective: reverse evolution. *Evolution* **55**, 653–660 (2001).
- Collin, R. & Cipriani, R. Dollo's law and the re-evolution of shell coiling. *Proc. R. Soc. Lond. B* **270**, 2551–2555 (2003).
- Domes, K., Norton, R. A., Maraun, M. & Scheu, S. Reevolution of sexuality breaks Dollo's law. *Proc. Natl Acad. Sci. USA* **104**, 7139–7144 (2007).
- Kohlsdorf, T. & Wagner, G. P. Evidence for the reversibility of digit loss: a phylogenetic study of limb evolution in *Bachia* (Gymnophthalmidae: Squamata). *Evolution* **60**, 1896–1912 (2006).
- Whiting, M. F., Bradler, S. & Maxwell, T. Loss and recovery of wings in stick insects. *Nature* **421**, 264–267 (2003).
- Chippindale, P. T., Bonett, R. M., Baldwin, A. S. & Wiens, J. J. Phylogenetic evidence for a major reversal of life-history evolution in plethodontid salamanders. *Evolution* **58**, 2809–2822 (2004).
- Goldberg, E. E. & Igic, B. On phylogenetic tests of irreversible evolution. *Evolution* **62**, 2727–2741 (2008).
- Collin, R. & Miglietta, M. P. Reversing opinions on Dollo's Law. *Trends Ecol. Evol.* **23**, 602–609 (2008).
- Wagner, G. P. The logical structure of irreversible systems transformations: a theorem concerning Dollo's law and chaotic movement. *J. Theor. Biol.* **96**, 337–346 (1982).
- Zufall, R. A. & Rausher, M. D. Genetic changes associated with floral adaptation restrict future evolutionary potential. *Nature* **428**, 847–850 (2004).
- Lewontin, R. C. Is nature probable or capricious? *Bioscience* **16**, 25–27 (1966).
- Page, M. Limpets break Dollo's Law. *Trends Ecol. Evol.* **19**, 278–280 (2004).
- Thornton, J. W. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nature Rev. Genet.* **5**, 366–375 (2004).
- Dean, A. M. & Thornton, J. W. Mechanistic approaches to the study of evolution: the functional synthesis. *Nature Rev. Genet.* **8**, 675–688 (2007).
- Yokoyama, S., Tada, T., Zhang, H. & Britt, L. Elucidation of phenotypic adaptations: molecular analyses of dim-light vision proteins in vertebrates. *Proc. Natl Acad. Sci. USA* **105**, 13480–13485 (2008).
- Bridgham, J. T., Carroll, S. M. & Thornton, J. W. Evolution of hormone-receptor complexity by molecular exploitation. *Science* **312**, 97–101 (2006).
- Ortlund, E. A., Bridgham, J. T., Redinbo, M. R. & Thornton, J. W. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **317**, 1544–1548 (2007).
- Wurtz, J. M. et al. A canonical structure for the ligand-binding domain of nuclear receptors. *Nature Struct. Biol.* **3**, 87–94 (1996).
- Smith, J. M. Natural selection and the concept of a protein space. *Nature* **225**, 563–564 (1970).
- Majerus, M. E. N. *Melanism: Evolution in Action* 151–154 (Oxford Univ. Press, 1998).
- Dawkins, R. *Blind Watchmaker* (W.W. Norton, 1994).
- Weinreich, D. M., Delaney, N. F., Depristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
- Dennett, D. C. *Darwin's Dangerous Idea: Evolution and the Meanings of Life* (Simon & Schuster, 1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Supported by National Science Foundation IOB-0546906, National Institutes of Health R01-GM081592 and F32-GM074398, and a Sloan Foundation Fellowship to J.W.T. We thank M. Harms and members of the Thornton, Cresko and Phillips laboratories for comments.

Author Contributions J.T.B. and J.W.T. conceived the experiments. J.T.B. performed the functional experiments, E.A.O. the structural analysis, and J.W.T. the phylogenetic analysis. J.T.B., E.A.O. and J.W.T. interpreted the results. J.T.B. and J.T. wrote the paper.

Author Information The atomic coordinates and structure factors for AncGR2 have been submitted to the Protein Data Bank (PDB) under accession number 3GN8. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.W.T. (joet@uoregon.edu).

METHODS

Ancestral protein sequences. AncGR1 and AncGR2 sequences were inferred by maximum likelihood³¹ using PAML 3.15 software on the maximum-likelihood phylogeny of 60 amino acid sequences of extant steroid and related receptors (see refs 23 and 24 for details). In brief, the likelihood of each possible amino acid state was calculated given the extant sequence data, the maximum-likelihood phylogeny, the Jones–Taylor–Thornton amino acid replacement model (which had 100% posterior probability in a Bayesian evaluation of several protein models), and a gamma distribution of among-site rate variation. The maximum-likelihood amino acid sequences of the LBDs (including the carboxy-terminal extension) of AncGR1 and AncGR2 were back-translated assuming human codon bias. Coding DNAs were then synthesized (GenScript), verified by sequencing, and cloned into pSG5-Gal4-DBD with the human glucocorticoid receptor hinge domain for expression and characterization. The functions of AncGR1- and AncGR2-LBD fusion proteins, assayed as described below, were robust to statistical ambiguity in the inferred ancestral sequence²⁴.

Molecular biology. The hormone-dependent transcriptional activity of resurrected ancestral receptors and their variants was assayed using a luciferase reporter system. CHO-K1 cells were grown in 96-well plates and transfected with 1 ng of receptor plasmid, 100 ng of a UAS-driven firefly luciferase reporter (pFRLuc), and 0.1 ng of the constitutive phRLtk *Renilla* luciferase reporter plasmid, using Lipofectamine and Plus Reagent in OPTIMEM (Invitrogen). After 4 h, transfection medium was replaced with phenol-red-free α MEM supplemented with 10% dextran-charcoal-stripped FBS (Hyclone). After overnight recovery, cells were incubated in triplicate with aldosterone, cortisol or 11-deoxycorticosterone from 10^{-11} to 10^{-5} M for 24 h, then assayed using Dual-Glo luciferase (Promega). Firefly luciferase activity was normalized by *Renilla* luciferase activity. Dose-response relationships were estimated using nonlinear regression in Prism4 software (GraphPad Software, Inc.); fold increase in activation was calculated relative to vehicle-only control. Mutagenesis to recapitulate historical substitutions was performed using QuikChange (Stratagene) and verified by sequencing.

Structural biology. The atomic structure of AncGR2-LBD was determined using X-ray crystallography. AncGR2-LBD cDNA (residues 1–248) was cloned into

pMALCH10T (a gift from J. Tesmer) and expressed as a maltose-binding protein/TEV-fusion protein in BL21(DE3) pLys cells in the presence of 50 μ M dexamethasone using standard methods. Expressed protein was purified using affinity chromatography. After TEV cleavage, the tagged fusion protein was removed using a nickel affinity column, polished by gel filtration, dialysed (200 mM sodium chloride, 50 μ M HEPES, pH 7.8 and 50 μ M CHAPS), and concentrated to 3–5 mg ml⁻¹. Crystals of AncGR2-LBD with dexamethasone were grown by hanging drop vapour diffusion at 22 °C from solutions containing 0.75 μ l of protein at 3–5 mg ml⁻¹ and 0.75 μ l⁻¹ of crystallant (0.5–0.75 M ammonium, pH 7.4), and a 21-amino-acid nuclear receptor box-3 peptide of glucocorticoid receptor coactivator human TIF2 (⁺H₃N-PVSPKKKENALLRYLLDKDDT-CO₂⁻, Synbiosci). Crystals were cryoprotected in crystallant with 25% glycerol and flash-cooled in liquid N₂. Data to 2.5 Å resolution were collected at 100K at the South East Regional Collaborative Access Team at the Advanced Photon Source, and were processed and scaled with HKL2000 (ref. 32; Supplementary Table 1). Initial phases for the AncGR2-cortisol complex were determined using PHASER³³ in the CCP4 software suite. The previously described homology model of AncGR2 (ref. 24) was used as a molecular replacement search model. All structures were refined using COOT³⁴ and CNS³⁵. The X-ray crystal structure of AncGR2 (PDB accession 3GN8) was compared to the model of AncGR1 (ref. 24), which was previously generated by homology modelling based on the X-ray crystal structure of its evolutionary precursor AncCR (PDB accession 2Q1V), with which it is identical at 90% of sites.

31. Yang, Z., Kumar, S. & Nei, M. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* **141**, 1641–1650 (1995).
32. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
33. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
34. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr.* **60**, 2126–2132 (2004).
35. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

LETTERS

Population context determines cell-to-cell variability in endocytosis and virus infection

Berend Snijder^{1,2}, Raphael Sacher^{1,2}, Pauli Rämö¹, Eva-Maria Damm¹, Prisca Liberali¹ & Lucas Pelkmans¹

Single-cell heterogeneity in cell populations arises from a combination of intrinsic and extrinsic factors^{1–3}. This heterogeneity has been measured for gene transcription, phosphorylation, cell morphology and drug perturbations, and used to explain various aspects of cellular physiology^{4–6}. In all cases, however, the causes of heterogeneity were not studied. Here we analyse, for the first time, the heterogeneous patterns of related cellular activities, namely virus infection, endocytosis and membrane lipid composition in adherent human cells. We reveal correlations with specific cellular states that are defined by the population context of a cell, and we derive probabilistic models that can explain and predict most cellular heterogeneity of these activities, solely on the basis of each cell's population context. We find that accounting for population-determined heterogeneity is essential for interpreting differences between the activity levels of cell populations. Finally, we reveal that synergy between two molecular components, focal adhesion kinase and the sphingolipid GM1, enhances the population-determined pattern of simian virus 40 (SV40) infection. Our findings provide an explanation for the origin of heterogeneity patterns of cellular activities in adherent cell populations.

Virus infection can be considered the end result of the concerted action of many cellular processes, including endocytosis and the establishment of membrane lipid composition^{7,8}. It is well accepted that these activities display heterogeneity within a population of cells, but the underlying causes of this heterogeneity are not known. We can therefore not predict how such activities behave in a population of cells⁹.

To study the heterogeneity of these activities, we examined large populations of single monoclonal adherent human cells after three days of growth. These were infected by rotavirus, dengue virus, mouse hepatitis virus (MHV) or SV40. MHV uses clathrin-mediated endocytosis (CME) for infectious entry into the cell¹⁰, and SV40 binds to the sphingolipid GM1 for host cell attachment and entry^{11,12}. To analyse CME, we allowed cells to internalize fluorescent transferrin¹³. To determine the amount of GM1, which is enriched in lipid rafts on the cell surface¹⁴, we exposed cells to fluorescent cholera toxin B¹⁵.

We observed that the cell-to-cell variation or heterogeneity of these activities displays specific patterns in both cancer- (HeLa and A431) (Fig. 1A) and non-cancer-derived (MCF10A) cells (see below). The patterns could not have been the result of physical constraints on the accessibility of virus particles or fluorescent probes to particular cells, because different types of pattern were observed.

A potential source of single-cell heterogeneity is the deterministic interplay between the phenotypic state of a cell and its activities^{2,16}, mediated by sensors and signal transduction networks. Growing isogenic adherent mammalian cells create cell islets and regions that are sparsely or densely populated, to which cells adapt their size, shape and rate of proliferation^{17,18} (Supplementary Movie 1), and display

several discrete subpopulations¹⁹. Together, this implies that patterns of varying cellular states can arise from the adaptation of individual human cells to the particular 'population context' in which they reside. The coupling between cell size and timing of meiosis in yeast²⁰, between cell size and the determination of phage infection outcome in *Escherichia coli*²¹, and the self-assembly of prokaryotic cells into complex colony patterns²², indicate that such mechanisms also operate in unicellular organisms. We therefore hypothesized that the heterogeneity patterns that we observed are caused by regulatory mechanisms that couple the phenotypic state of a cell to intracellular processes on which virus infection depends, such as endocytosis and the lipid composition of the cell surface.

To reveal the presence of such mechanisms, we used three data-driven modelling approaches on several single-cell measurements obtained by computerized image analysis and supervised machine learning from multiple large populations of cells. (See Fig. 1B for an overview of the methods used, and Supplementary Methods for detailed information.) We first measured for each cell the size of the population to which it belonged, its local cell density, its position on a cell islet edge, its cell size, its mitotic state and its apoptotic state (microenvironment and cell state). By delineating how these parameters interact with each other, we quantified how the characteristics of a cell population are determined. This showed that each parameter, in particular population size, local cell density, position on a cell islet edge, and cell size, represents a population-determined property of an individual cell. The same interactions were observed in cancer cells (Supplementary Fig. 1), normal diploid cells (see below) and primary cells (not shown), despite their different morphologies and growth rates.

For each cell, we next determined whether it was infected, or we quantified its CME activity or cell-surface level of GM1 (cellular activities). We then used the population-determined properties as predictors to model the probability of infection, the activity of CME and the amount of GM1 on the cell surface. In each instance, models could be derived that had good fits (Fig. 1C and Supplementary Fig. 2). The model parameters demonstrated the existence of extensive and specific regulatory mechanisms between population context and virus infection, endocytosis and membrane lipid composition (Fig. 1D). Importantly, the models were able to predict accurately the heterogeneity patterns of cellular activities, solely based on a quantitative assessment of the population context and the state of individual cells (Fig. 1E).

An interesting question that arises is to what extent these population-determined effects contribute to the total variation observed in a population. Strikingly, variance analysis (Supplementary Methods) demonstrated that variation determined by population context is a major component of the total variation observed. For rotavirus, SV40 and dengue virus infection, our measures of the local environment

¹Institute of Molecular Systems Biology, ETH Zurich (Swiss Federal Institute of Technology), Wolfgang Pauli-Strasse 16, CH-8093 Zurich, Switzerland. ²Zurich PhD Program in Molecular Life Sciences, Zurich, Switzerland.

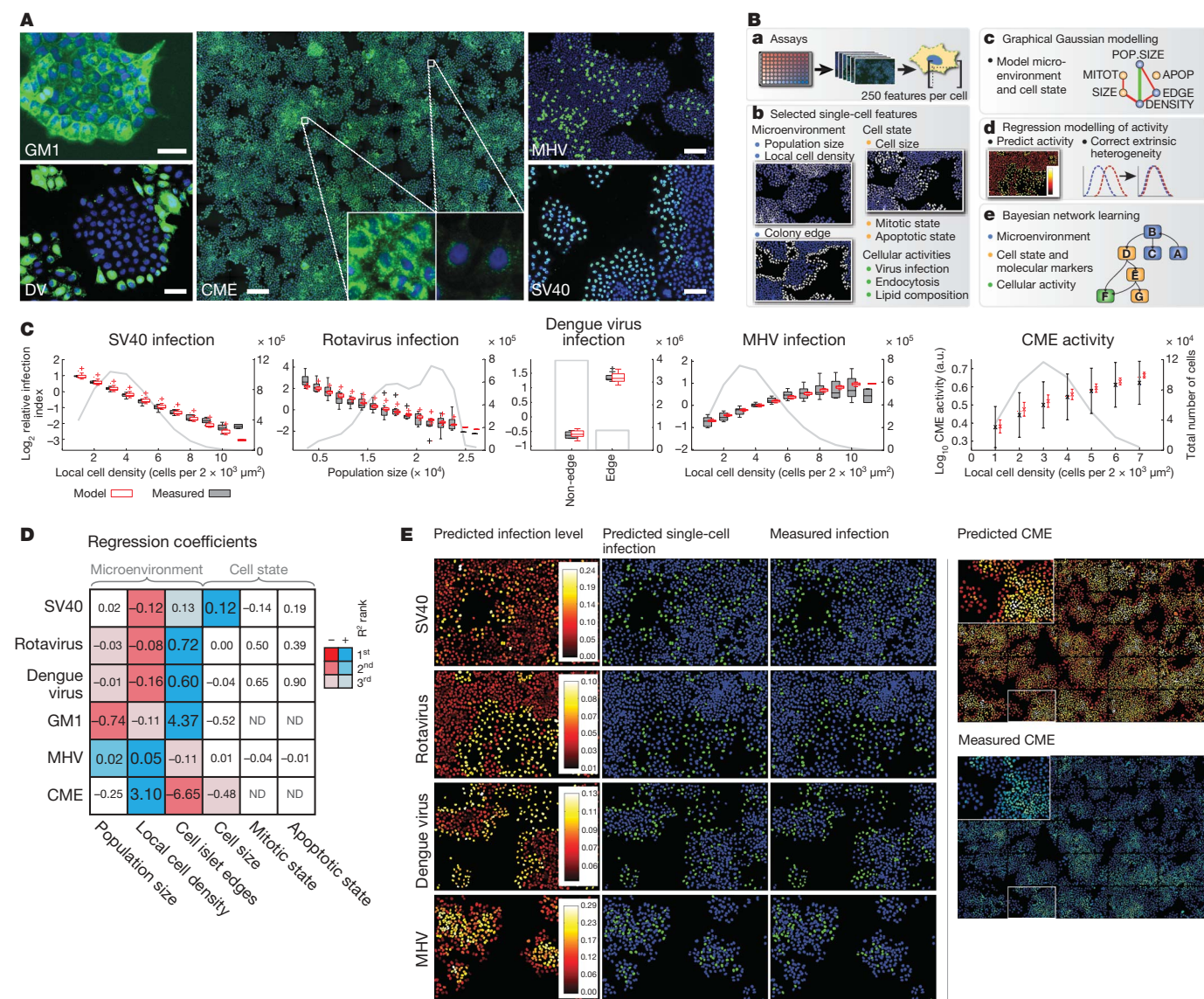


Figure 1 | Modelling and predicting diverse activity patterns in heterogeneous human cell populations. **A**, GM1 levels on the plasma membrane of A431 cells and dengue virus (DV) infection in HeLa cells are higher in cells at islet edges (upper and lower left panels). MHV infection and CME (upper right and middle panels) are highest in HeLa cells in crowded regions. SV40 (lower right panel) infects HeLa cells that are large and spread out. Scale bars: GM1 and dengue virus, 50 μm ; CME, MHV and SV40 respectively 300, 250 and 150 μm . **B**, Overview of the experimental, computational and statistical methods used in this study (see Methods and Supplementary Methods for details). **C**, Virus infection ($n = 9$; grey

box-plots) and CME (black lines; average \pm single-cell s.d.) plotted against population-context parameters. Model fits are depicted in red. Right axes display the number of cells measured (light-grey bars and lines).

D, Regression coefficients between population-context parameters (columns) and cellular activities (rows), colour-coded according to coefficient sign and top-three ranked explanatory power (see legend). ND, not determined. **E**, Model-predicted virus infection and CME patterns displayed on computer-generated reproductions of original images (left panels). Predicted and measured single-cell activity levels are shown (middle and right panels).

and cellular state account for at least 60%, 74% and 82% of the respective total variation for each activity (Fig. 2a). For CME and GM1 content on the cell surface, variation determined by population context appeared to constitute at least 30% and 21% of the respective total variation (Fig. 2a).

To understand better the variation unexplained by population context, we analysed the variance in CME activity in groups of cells with similar population-determined phenotypic properties. This revealed that cells in certain population contexts displayed as much as a 20-fold lower variation than randomly selected cells (inset Fig. 2a). When we analysed this variance relative to the mean CME activity in these groups of cells (coefficient of variation), we observed a threefold lower coefficient of variation in cells with a high local cell density (Fig. 2b). This indicates that these cells have, besides a high CME activity, mechanisms in place that exert a tight control on this activity.

The finding that much of the variation in virus infection, endocytosis and lipid composition of the cell surface is deterministically established by the adaptation of cells to their population context reveals a fundamental problem in our current methods of studying differences in these activities between cell populations. We illustrate this by comparing the fraction of cells infected with dengue virus (Fig. 2c) and the single-cell distributions of CME activity (Fig. 2d) between unperturbed populations of different sizes. The activities display levels that are statistically different (respectively $P = 2.4 \times 10^{-6}$ and $P < 10^{-10}$) between the different unperturbed populations. However, after normalization of both activities with activities predicted by their respective models, no significant differences were found (P values of respectively 0.97 and 0.75). This is the correct conclusion, as both populations were unperturbed and only differed in their population context. A similar problem exists with supervised and unsupervised clustering methods. Even though such methods correctly identify different phenotypic

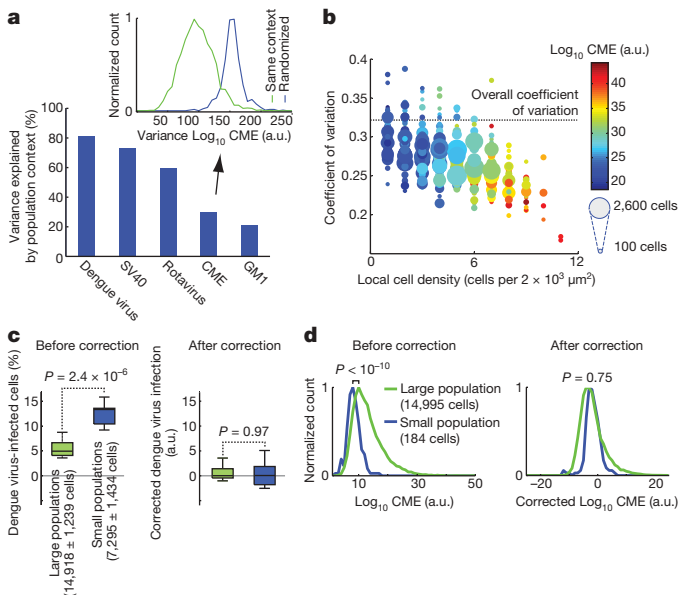


Figure 2 | Most variation observed in cellular activities is population determined and must be accounted for. **a**, Population-determined variance (percentage of total) in the levels of virus infection, CME and GM1 content (Supplementary Methods). Variance of CME in groups of cells in similar population contexts (insert; green) compared with randomized groups of cells (blue). **b**, Coefficient of variation (s.d./mean) of CME in groups of cells with similar population contexts plotted against local cell density ($r^2 = -0.57$, $P < 10^{-10}$). Number of cells per group (dot size) and average activity level (colour) are indicated (see legend). **c**, Significantly different dengue virus infection ($P = 2.4 \times 10^{-6}$) between unperturbed populations of cells with different population properties ($n = 9$). Model correction abolishes this difference ($P = 0.97$). **d**, Significantly different single-cell CME activities ($P < 10^{-10}$) between unperturbed populations of cells with different population properties. Model correction abolishes this difference ($P = 0.75$).

states (Supplementary Fig. 3), they will not reveal whether these phenotypic states are determined by cellular adaptations to the population context. Therefore, these methods will be unable to distinguish between a change in cellular activity that is a consequence of an altered population context and a change that is a consequence of a direct perturbation.

Besides these fundamental implications, our analysis reveals suggestions for the mechanisms underlying population-determined heterogeneity of virus infection (see ranking in Fig. 1d and Supplementary Fig. 4). The infection efficiency of rotavirus is strongly increased in cells that grow in sparsely populated areas. Indeed, rotavirus uses integrins as its receptor²³, which are probably highly expressed on the surface of cells that grow sparsely²⁴. MHV, on the contrary, prefers to infect cells that grow at high local density, similar to where CME is most active. Dengue virus infection occurs almost exclusively in cells located on the edge of cell islets, indicating that it relies on mechanisms highly active in polarized cells²⁵.

The heterogeneity signature of SV40 infection demonstrated a particular preference for large and spread-out cells in both cancer- and non-cancer-derived cell lines (see below). As these cells are predominantly located on the edge of cell islets and in areas that are sparsely populated, they also have higher amounts of GM1 on the cell surface (Fig. 1D), the receptor for SV40^{11,12}. In addition, we previously found that focal adhesion kinase (FAK or PTK2) is required for SV40 infection⁸, and regulatory links between the micro-environment of cells and FAK²⁴, and between FAK and sphingolipid membrane domains, have been reported^{26,27}.

To test whether FAK and GM1 can be placed in a causal network that determines the heterogeneity pattern of SV40 infection, we applied Bayesian network learning⁵ (Supplementary Methods) on

single-cell measurements of SV40 infection, amount of GM1 on the cell surface, level of Y397-phosphorylated (activated) FAK and population-determined phenotypic properties (Fig. 3a and Supplementary Fig. 5). We chose to use MCF10A cells to point out the relevance of these phenomena also for non-cancer-derived, normal diploid cells. This identified a unique causal network that combines single-cell microenvironmental parameters with molecular components and virus infection (Fig. 3b). We found that cell density is coupled to the regulation of cell size through GM1 cell surface content, and FAK activation. The last two are also determinants of SV40 infection, in addition to as yet unidentified factors that are regulated by cell density. We validated three causal interactions in the Bayesian network. Exogenous addition of GM1 to MCF10A cells increased the levels of phosphorylated FAK and increased SV40 infection. Additionally, knockdown of FAK by RNA interference strongly decreased SV40 infection (Supplementary Fig. 7). Thus, by applying quantitative data-driven modelling on a combination of micro-environmental, cellular and molecular parameters, we reveal mechanistic insights into the establishment of cellular heterogeneity.

We next wondered why the heterogeneity of SV40 infection is more deterministic (74% explained) than the GM1 levels and FAK activity on which it depends (respectively 30% and 65% explained) (Fig. 3b). Interestingly, when we validated the causal interaction

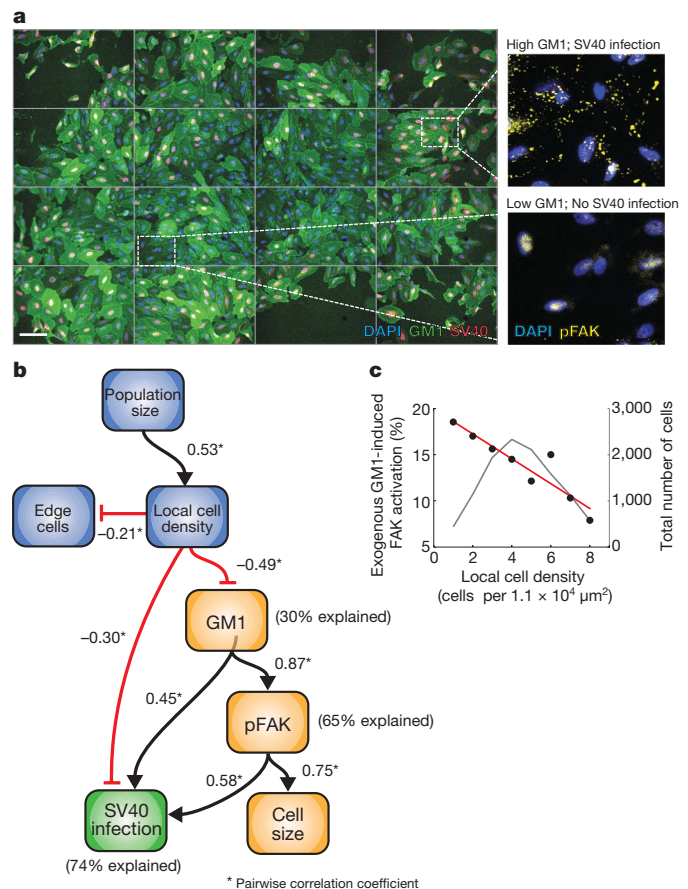


Figure 3 | SV40 infection and cell size are co-regulated by GM1 and active FAK, depending on the local cell density. **a**, SV40 infection (red), GM1 content (green) and Y397-phosphorylated FAK (enlargements; yellow) in MCF10A cells. Scale bar, 150 μm . **b**, Bootstrapped Bayesian network learning (Supplementary Fig. 6 and Supplementary Methods) reveals causal links between population-context parameters, GM1 content, pFAK levels and SV40 infection. Edges are drawn according to the sign of pairwise single-cell correlations. Variation in levels of GM1, pFAK and SV40 infection explained by population context are respectively 30%, 65% and 74%. **c**, Scatter plot of exogenous GM1-induced increase in pFAK (percentage of non-induced) against local cell density. Linear fit (red) and total number of cells for each measurement (right axis, grey) are shown.

between GM1 and FAK by exogenous addition of GM1, we observed a 2.5-fold stronger activation of FAK in sparse cells compared with dense cells (Fig. 3c). This indicates that GM1 and FAK synergize in sparse cells, leading to the more deterministic downstream activity of SV40 infection. Such synergistic mechanisms can partly cancel out the intrinsic and uncorrelated noise of individual components, increasing the deterministic nature of complex activities at the cellular level^{2,28}.

We have shown that much of the variation in virus infection, endocytosis and membrane lipid composition is determined by the adaptation of cells to their population context, and uncovered basic predictive principles by which non-differentiated cells create complex patterns of activity. Similar mechanisms may determine the heterogeneity of other cellular processes. Perturbation screens, combined with quantitative modelling of single cells in their population context, will further reveal the molecular networks that regulate heterogeneity patterns in cell populations. The principles described here most likely operate in all systems of collective cellular behaviour, from prokaryotic colonies to multicellular organisms.

METHODS SUMMARY

All human cell lines were maintained under standard tissue culture conditions. All assays were performed in 96-well plates, and cells were imaged on automated widefield cellWoRx microscopes (Applied Precision) or ImageXpress Micro microscopes (Molecular Devices) with $\times 10$ or $\times 20$ magnification. Infection assays were performed as described (Supplementary Information). CME was measured using Alexa Fluor 488-conjugated transferrin, and GM1 levels were visualized using cholera toxin B conjugated to Alexa Fluor 568 as described⁸. Phosphorylated-FAK [Y397] was visualized by using standard indirect immunofluorescence protocols. High-content single-cell image analysis was performed using CellProfiler²⁹ and additional image analysis algorithms written specifically for this study (Supplementary Information). In short, cells were identified by object detection on DAPI images, and these regions were typically expanded to cover the cytoplasm. Shape features and features describing the texture and intensities in all channels were extracted for these regions of interest. Support vector machine learning was applied (P. R  m  , R. Sacher, B. Snijder, B. Begemann and L. Pelkmans, submitted) for the classification of diverse cellular phenotypes, including interphase, mitotic and apoptotic cellular states, as well as infectious phenotypes and technical artefacts. Globally, these data were further analysed using several statistical methods. First, graphical Gaussian modelling quantified how a population shapes the property distributions of individual cells and how individual properties influence each other, which was experimentally validated by live cell imaging. Second, probit regression modelling revealed how virus infection depends on population-determined properties of individual cells, as did weighted linear regression for GM1 content and CME. Finally, bootstrapped Bayesian network learning was applied to identify causal interactions between properties of the microenvironment, FAK-phosphorylation, GM1 content and SV40 infection in MCF10A cells. Selected inferred causal links were experimentally validated by using GM1 addback and short interfering RNA (siRNA)-mediated silencing of FAK as described⁸. A detailed description of the methods performed in this study is provided in the Supplementary Information.

Received 20 March; accepted 10 July 2009.

Published online 26 August 2009.

1. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
2. Maheshri, N. & O'Shea, E. K. Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 413–434 (2007).
3. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–226 (2008).
4. Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444**, 643–646 (2006).
5. Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D. A. & Nolan, G. P. Causal protein-signaling networks derived from multiparameter single-cell data. *Science* **308**, 523–529 (2005).

6. Keren, K. *et al.* Mechanism of shape determination in motile cells. *Nature* **453**, 475–480 (2008).
7. Liberali, P., Ramo, P. & Pelkmans, L. Protein kinases: starting a molecular systems view of endocytosis. *Annu. Rev. Cell Dev. Biol.* **24**, 501–523 (2008).
8. Pelkmans, L. *et al.* Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis. *Nature* **436**, 78–86 (2005).
9. Wilkinson, D. J. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Rev. Genet.* **10**, 122–133 (2009).
10. Eifart, P. *et al.* Role of endocytosis and low pH in murine hepatitis virus strain A59 cell entry. *J. Virol.* **81**, 10758–10768 (2007).
11. Neu, U., Woellner, K., Gauglitz, G. & Stehle, T. Structural basis of GM1 ganglioside recognition by simian virus 40. *Proc. Natl Acad. Sci. USA* **105**, 5219–5224 (2008).
12. Pelkmans, L. Secrets of caveolae- and lipid raft-mediated endocytosis revealed by mammalian viruses. *Biochim. Biophys. Acta* **1746**, 295–304 (2005).
13. Conner, S. D. & Schmid, S. L. Regulated portals of entry into the cell. *Nature* **422**, 37–44 (2003).
14. Mayor, S. & Pagano, R. E. Pathways of clathrin-independent endocytosis. *Nature Rev. Mol. Cell Biol.* **8**, 603–612 (2007).
15. Holmgren, J., Lonnroth, I. & Svennerholm, L. Tissue receptor for cholera exotoxin: postulated structure from studies with GM1 ganglioside and related glycolipids. *Infect. Immun.* **8**, 208–214 (1973).
16. Sacher, R., Stergiou, L. & Pelkmans, L. Lessons from genetics: interpreting complex phenotypes in RNAi screens. *Curr. Opin. Cell Biol.* **20**, 483–489 (2008).
17. Eagle, H. & Levine, E. M. Growth regulatory effects of cellular interaction. *Nature* **213**, 1102–1106 (1967).
18. Castor, L. N. Flattening, movement and control of division of epithelial-like cells. *J. Cell. Physiol.* **75**, 57–64 (1970).
19. Slack, M. D., Martinez, E. D., Wu, L. F. & Altschuler, S. J. Characterizing heterogeneous cellular responses to perturbations. *Proc. Natl Acad. Sci. USA* **105**, 19306–19311 (2008).
20. Nachman, I., Regev, A. & Ramanathan, S. Dissecting timing variability in yeast meiosis. *Cell* **131**, 544–556 (2007).
21. St-Pierre, F. & Endy, D. Determination of cell fate selection during phage lambda infection. *Proc. Natl Acad. Sci. USA* **105**, 20705–20710 (2008).
22. Ben-Jacob, E., Cohen, I. & Gutnick, D. L. Cooperative organization of bacterial colonies: from genotype to morphotype. *Annu. Rev. Microbiol.* **52**, 779–806 (1998).
23. Lopez, S. & Arias, C. F. Multistep entry of rotavirus into cells: a Versaillesque dance. *Trends Microbiol.* **12**, 271–278 (2004).
24. Geiger, B., Spatz, J. P. & Bershadsky, A. D. Environmental sensing through focal adhesions. *Nature Rev. Mol. Cell Biol.* **10**, 21–33 (2009).
25. Neumann, A. K., Thompson, N. L. & Jacobson, K. Distribution and lateral mobility of DC-SIGN on immature dendritic cells—implications for pathogen uptake. *J. Cell Sci.* **121**, 634–643 (2008).
26. Iwabuchi, K. *et al.* Reconstitution of membranes simulating “glycosignaling domain” and their susceptibility to Lyso-GM3. *J. Biol. Chem.* **275**, 15174–15181 (2000).
27. Palazzo, A. F., Eng, C. H., Schlaepfer, D. D., Marcantonio, E. E. & Gundersen, G. G. Localized stabilization of microtubules by integrin- and FAK-facilitated Rho signaling. *Science* **303**, 836–839 (2004).
28. Newman, J. R. & Weissman, J. S. Systems biology: many things from one. *Nature* **444**, 561–562 (2006).
29. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge H. Verheije and L. Burleigh for providing images of MHV and dengue virus infection, G. Jurisic for providing primary cells and help with experiments, and all members of the laboratory for comments on the manuscript. P.R. is supported by the European Molecular Biology Organisation and the Human Frontiers Science Program, E.-M.D. by Onco Suisse and P.L. by the Federation of European Biochemical Societies. L.P. is supported by the ETH Z  rich, SystemsX.ch, the Swiss National Science Foundation and the European Union.

Author Contributions L.P. supervised and conceived the project. R.S., B.S., E.-M.D. and P.L. performed experiments, B.S. and P.R. developed computational image analysis methods, B.S. performed all computational image analysis, B.S. and P.R. conceived the statistical analysis methods, B.S. performed all statistical analysis, L.P. and B.S. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to L.P. (pelkmans@imsb.biol.ethz.ch).

LETTERS

Integration of neuronal clones in the radial cortical columns by EphA and ephrin-A signalling

Masaaki Torii^{1,2}, Kazue Hashimoto-Torii¹, Pat Levitt² & Pasko Rakic¹

The cerebral cortex is a laminated sheet of neurons composed of the arrays of intersecting radial columns^{1–3}. During development, excitatory projection neurons originating from the proliferative units at the ventricular surface of the embryonic cerebral vesicles migrate along elongated radial glial fibres⁴ to form a cellular infrastructure of radial (vertical) ontogenetic columns in the overlying cortical plate⁵. However, a subpopulation of these clonally related neurons also undergoes a short lateral shift and transfers from their parental to the neighbouring radial glial fibres⁶, and intermixes with neurons originating from neighbouring proliferative units^{5,7}. This columnar organization acts as the primary information processing unit in the cortex^{1,8,9}. The molecular mechanisms, role and significance of this lateral dispersion for cortical development are not understood. Here we show that an Eph receptor A (EphA) and ephrin A (EfnA) signalling-dependent shift in the allocation of clonally related neurons is essential for the proper assembly of cortical columns. In contrast to the relatively uniform labelling of the developing cortical plate by various molecular markers and retrograde tracers in wild-type mice, we found alternating labelling of columnar compartments in *EfnA* knockout mice that are caused by impaired lateral dispersion of migrating neurons rather than by altered cell production or death. Furthermore, *in utero* electroporation showed that lateral dispersion depends on the expression levels of EphAs and ephrin-As during neuronal migration. This so far unrecognized mechanism for lateral neuronal dispersion seems to be essential for the proper intermixing of neuronal types in the cortical columns, which, when disrupted, might contribute to neuropsychiatric disorders associated with abnormal columnar organization^{8,10}.

Functional columns in the neocortex emerge developmentally from a proliferative zone protomap, and consist of several classes of interconnected neurons^{1,11,12}. In the present study, we have explored the possible roles of A-type Eph receptors and their ligands, ephrin-As, in the integration of neurons in radial columns because they have been implicated in controlling cell positioning in a variety of developmental contexts by sorting cell types^{13–15}, restricting their intermingling¹⁶, or regulating their migration¹⁷. First, to determine their expression during the period when most postmitotic neurons migrate radially to the cortical plate, sections were incubated with human Fc-tagged ephrin-A (EFNA) and EPH receptor A (EPHA) proteins that bind to all EphAs and ephrin-As, respectively. We observed an elaborate but stereotypical binding pattern of both receptors and ligands across the thickness of the neocortex with a peak expression in the intermediate zone, where they exhibited opposing gradients (Supplementary Fig. 1). In fact, several *Epha* receptors and *EfnA* ligands were detected in the subventricular zone (SVZ) and/or the intermediate zone, through which neurons that originated in the ventricular zone migrate to the cortical plate (Supplementary Fig. 1).

To examine the roles of EphA receptors and ephrin-A ligands in neuronal migration in the neocortex, we analysed the triple knockout (TKO) mouse¹⁸ for *EfnA1*, *EfnA3* and *EfnA5* that account for almost all ephrin-A genes in the developing neocortex¹⁹. To minimize the effects of mistargeted afferent and efferent projections^{19–21}, we performed most analyses at postnatal day 0 (P0) or embryonic stages before the establishment of these projections. It has been demonstrated that EphA and ephrin-A signalling have a proapoptotic effect in the embryonic cortex without affecting the proliferation or cell-cycle progression²². Consistently, proliferation and cell-cycle progression were not affected in the TKO cortex (data not shown). In addition, similar to the *Epha7* knockout mice²², less than 10% of the 52 embryos showed an exencephalic phenotype at embryonic day (E)14.5, which is probably caused by decreased apoptosis of neural progenitors²². These animals were excluded from our analyses. The remaining TKO embryos did not show any difference in the number of apoptotic cells in the cortex, gross brain shape, brain mass, cortical surface size and cell numbers in a cortical hemisphere compared to wild-type embryos (Supplementary Fig. 2).

The overall patterning of cortical areas at P0, revealed by *in situ* hybridization, appeared unaffected in TKO (Supplementary Figs 2 and 3). However, more detailed analysis showed spotted, irregular distribution of cells labelled by these markers, and alternating thickening and thinning of the cortical plate in TKO mice (Supplementary Fig. 3). Analysis of the wild-type ($n \geq 7$) and TKO ($n \geq 9$) brains showed highly uneven distribution of CTIP2 (also known as Bcl11b)-labelled neurons in layer 5 along the tangential axis in the TKO, although their radial positioning was unaffected (Fig. 1a). Most domains with an excess of CTIP2⁺ neurons were found adjacent to domains with reduced CTIP2⁺ neurons (Fig. 1b, c). Similarly, Sox5- and Cutl1 (also known as Cux1)-labelled neurons in layers 5–6 and 2–4, respectively, showed normal radial but irregular tangential distribution of labelled neurons in the TKO (Fig. 1d, e). Notably, double labelling with CTIP2 and Cutl1, and with CTIP2 and Lmo4 (which is expressed in neurons of layers 2–6) showed that the irregular tangential distribution of marker-positive and -negative neuronal subtypes has a certain independency among cortical layers (Fig. 1e, f). Labelling of cell nuclei showed uneven cortical plate thickness and cell packing among neighbouring columnar domains (Fig. 1f and Supplementary Fig. 4). However, their increase and decrease, and those of marker-positive neurons, was not parallel, so that the thicker cortical domains with higher cell density often contained less Lmo4⁺ neurons in the upper cortical layers (Fig. 1f).

To gain insight into the functional organization of the TKO neocortex, corticotectal projection neurons (CTPN) and callosal projection neurons (CPN) were retrogradely labelled at P2 and observed at P4 ($n = 6$ per genotype). Whereas layer distribution of labelled neurons was normal, their numbers were highly variable along the tangential

¹Department of Neurobiology and Kavli Institute for Neuroscience, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA. ²Zilkha Neurogenetic Institute and Department of Cell and Neurobiology, Keck School of Medicine of USC, Los Angeles, California 90089, USA.

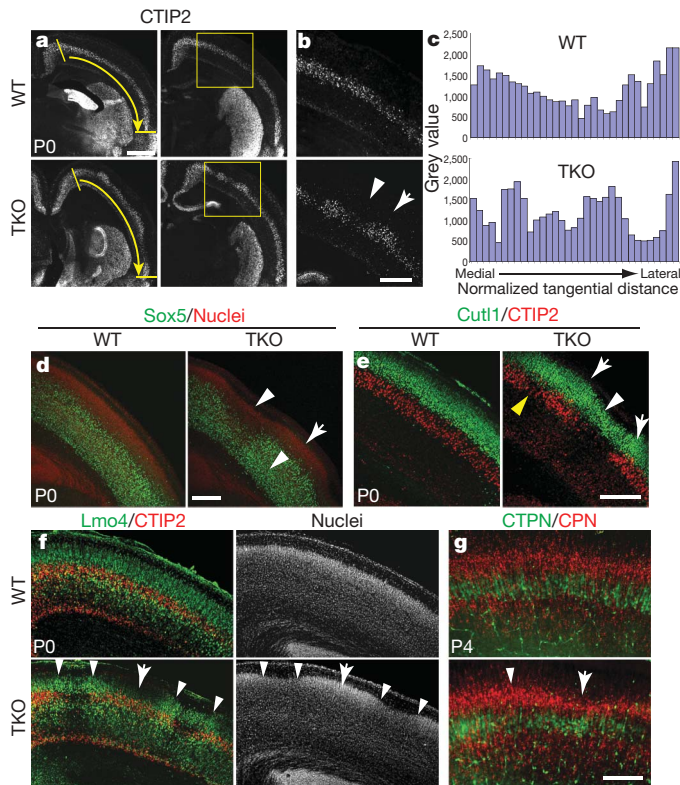


Figure 1 | Abnormal tangential organization of the cortical plate in the TKO neocortex. **a**, CTIP2 immunohistochemistry on wild-type (WT) and TKO brains. **b**, Higher magnification images of the boxed area in **a**. The arrow and arrowhead indicate the domains where abnormally more and less neurons are labelled, respectively. **c**, Quantification of the grey values of fluorescent CTIP2 labelling from the left panels in **a** along the medial-lateral axis (yellow arrows). Total labelling was similar in wild-type and TKO (39,361 and 39,036 in pixels, respectively). **d**, Irregular tangential distribution of Sox5⁺ neurons, associated with abnormal thickening and thinning of the cortical plate of TKO (arrow and arrowheads). **e**, Cutl1 and CTIP2 labelling shows an increase and decrease of labelled neuronal subpopulations (white arrows and arrowhead, respectively) in adjacent cortical domains in TKO. The increase and decrease was not always synchronized (yellow arrowhead). **f**, Lmo4 and CTIP2 staining shows distinct patterns of increased/decreased labelled neurons in TKO (arrow and arrowheads). **g**, The numbers of labelled CTPN and CPN were both highly variable among neighbouring columnar domains in TKO (arrow and arrowhead). Scale bars, 500 μ m (**a**) and 200 μ m (**b–g**).

axis in the TKO neocortex (Fig. 1g). Tangential axis-specific disruption was further confirmed at P14 by combining Cutl1 immunohistochemistry with 5-bromodeoxyuridine (BrdU) injection made at E14.5 (Supplementary Fig. 4). Furthermore, no apparent defect was observed in the organization of the marginal zone and pial surface, Reelin expression in Cajal–Retzius cells, and the organization of radial glial fibres in the TKO neocortex at P0 (Supplementary Fig. 5).

To test the possibility that neurons in TKO mice might be allocated into unusual cortical columns, we examined the positional relationship among the clonally related cortical neurons by infecting precursors with a green fluorescent protein (GFP) retrovirus at E12.5 and analysing the location of GFP⁺ neurons at P0. The results show that the width of ontogenetic columns of GFP⁺ clones in TKO was smaller (Fig. 2a, $n = 7$) than in the wild-type neocortex (Fig. 2a, $n = 10$). Furthermore, in most TKO cortical plate ($n = 5$ out of 7), the labelled neurons were aligned in straight columns with one- to two-cell diameters, which was never encountered in the wild-type cortical plate (Fig. 2b). This straight alignment of neurons was not due to a crowding effect of increased cell packing density (Supplementary Fig. 6). Because lateral movement of radially migrating neurons occurs within the SVZ and the intermediate zone during

the ‘multipolar stage’²³, we measured the lateral distance between migrating neurons and the parental radial fibres at the top of the intermediate zone at E14.5 in the neocortex infected with GFP retrovirus at E12.5. The distance was significantly less in TKO than that in wild-type (Fig. 2c), indicating that EphA and ephrin-A signalling are required for lateral dispersion/intermingling of neuronal clones during migration.

To address the mechanism of lateral dispersion, we performed cell ablation experiments, in which the diphtheria-toxin-A-chain fragment (DTA) and an enhanced yellow fluorescent protein (EYFP) reporter plasmids were electroporated into the cortex at E13.5, and the effects on the migrating neurons and upper layer neurons were examined at E15.5 and P0. As DTA causes inhibition of protein synthesis followed by apoptotic cell death²⁴, we observed death of infected progenitor cells in the ventricular zone and, importantly, also of the migrating neurons in the intermediate zone before they arrive in the cortical plate (Fig. 2d and also see Methods). With this system, we found that despite significant numbers of dying cells in the DTA-electroporated domains, surviving NeuroD⁺ young neurons²⁵ in the lower intermediate zone were distributed continuously along the tangential axis of the cortex in the wild-type neocortex at E15.5, irrespective of the size of electroporated areas (Fig. 2d, $n = 12$). In contrast, most electroporated TKO cortices showed interrupted distribution of NeuroD⁺ neurons (Fig. 2d, $n = 8$ out of 9), even in the brains that have very limited electroporated areas (Fig. 2e). Consistent with these observations, Cutl1-labelled upper layers were thinner in the TKO cortex at P0, specifically within the domains of DTA electroporation (Fig. 2f). Finally, we observed no difference in proliferation and neuronal production in the progenitors in the electroporated domains between wild-type and TKO (data not shown). These results demonstrate that the developing neocortex possesses a striking capacity to fill in the vacated space of ablated neurons with neighbouring migrating neurons, but that this capacity is lost in the TKO neocortex.

To examine the role of EphA and ephrin-A signalling in the lateral dispersion of neocortical neurons further, we co-electroporated either a control or an EphA7 expression plasmid with an EYFP reporter plasmid at E12.5, and examined the distribution of EYFP-labelled neurons at E18.5. In contrast to the random distribution of EYFP⁺ neurons in control brains (Fig. 3a, $n \geq 20$), EphA7-overexpressed neurons formed several EYFP⁺ columns (Fig. 3a, $n \geq 20$) in all cortical regions examined (Supplementary Fig. 7). Introducing another EphA receptor, EphA4, generated the same phenotype (data not shown). The exaggerated columnar pattern of EYFP⁺ neurons appears 1 day after EphA7 electroporation, and continues during embryonic and postnatal stages (Fig. 3a and Supplementary Fig. 7; $n = 7$ (E13.5) and $n \geq 20$ (P4)). Non-labelled neurons situated between EYFP⁺ columns had a similar cell packing density (Fig. 3a and Supplementary Figs 7 and 8) and the overall cytoarchitecture in the cortical plate was unaffected as previously reported²¹. Organization of apical dendrites and formation of radial glial fibres also were not affected grossly by EphA7 overexpression (Supplementary Fig. 8). TdT-mediated dUTP nick end labelling (TUNEL) and BrdU labelling at E18.5 after EphA7 electroporation showed no difference between EYFP⁺ and EYFP[−] domains (Supplementary Fig. 8). Furthermore, the cell-cycle index and cell death rate showed no change in the EphA7 electroporated neocortex (Supplementary Fig. 8). Finally, sequential electroporation with the EphA7 expression plasmid, the EYFP (first) and DsRed2 (second) marker plasmids, demonstrated that the isolated columns of EphA7-electroporated neurons consist of several clones (Supplementary Fig. 7). EphA7 electroporation into the TKO cortex did not form the columnar segregation (Supplementary Fig. 8). Together, these results indicate that EYFP⁺ columns are not formed by increased production or death of a subset of EYFP⁺ clones, but rather, by tangential segregation of EYFP⁺ and EYFP[−] neurons.

Because EphAs and ephrin-As can act as both receptors and ligands to mediate forward and reverse signalling²⁶, and both are required for

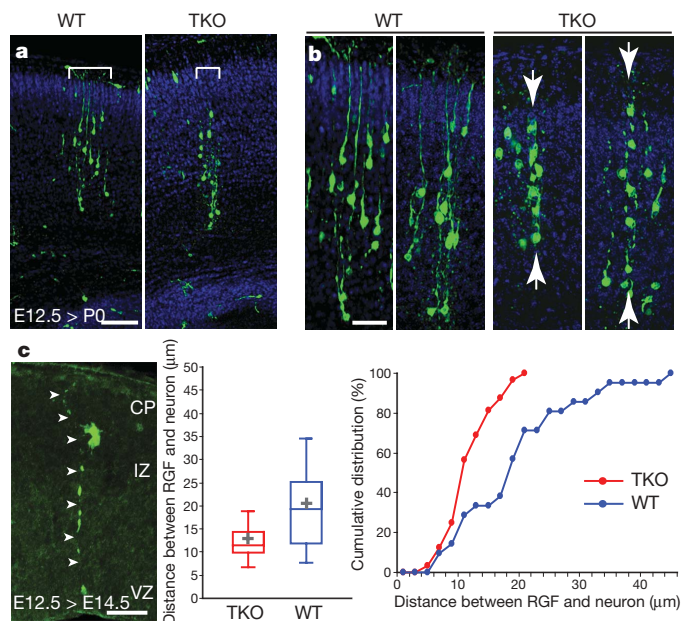
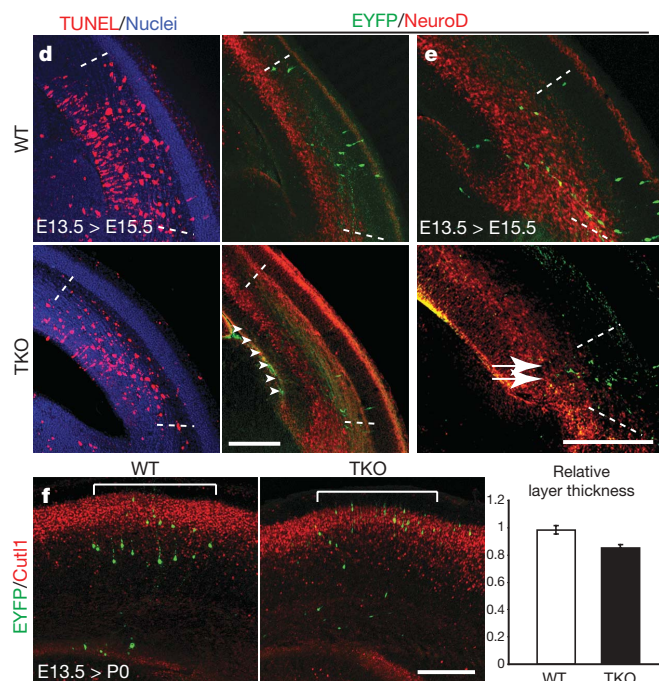


Figure 2 | Impaired lateral dispersion of cortical neurons in TKO. **a**, GFP-labelled (green) clonal neurons. The column width of the clonal neurons (brackets) was smaller in TKO. Nuclei were counterstained (blue). **b**, Magnified view of clonally related neurons. The right panel for wild-type shows the most extreme columnar case in wild-type. In the TKO cortex, straight alignments of GFP-labelled neurons (between arrows) were found. **c**, Left: representative image of GFP-labelled neurons along their parental radial glial fibre (RGF; arrowheads). CP, cortical plate; IZ, intermediate zone; VZ, ventricular zone. Middle: box-plot of the distance between labelled migrating neurons and the parental RGF. The box represents the median and the 25th and 75th percentiles. The top and bottom whiskers denote the 5th and 95th percentiles. Plus signs denote the mean distances. The distance was significantly shorter in TKO. $P = 0.0014$ (Mann–Whitney U test, $n = 32$ from 14 TKO brains, and 21 from 10 wild-type brains). Right: cumulative percentage plot from the same data. $P < 0.001$ (Kolmogorov–Smirnov test).

cell sorting^{13,15}, we electroporated ephrin-A2, -A5, and the truncated forms of EphA7 and EphA4, together with the EYFP plasmid to determine which signalling is involved in columnar segregation. We found segregation of EYFP⁺ and EYFP[−] neurons, but no effect on cell-cycle progression and cell death, in the ephrin-A2- and ephrin-A5-electroporated cortex (Supplementary Fig. 9, and data not shown). However, electroporation with the truncated EphAs, which can induce only reverse signalling as ligands, failed to induce segregation (Supplementary Fig. 9). These results indicate that forward signalling is necessary for tangential segregation, although we cannot rule out completely that reverse signalling might also be involved cooperatively.

To determine the mechanism of columnar segregation during radial migration, we analysed the E15.5 neocortex electroporated at E12.5. No difference was observed in the random probability of gene transfer among progenitors in the ventricular zone between control ($n = 16$) and EphA7-electroporated brains ($n = 23$) (Fig. 3b, c). In both groups, EYFP⁺ cells in the ventricular zone and cortical plate have radially elongated shapes without any evidence of tangential movement (Fig. 3a, b). In the intermediate zone of all cases, most of the EYFP⁺ migrating neurons assumed a multipolar shape. They were more randomly distributed than EYFP⁺ progenitors in the ventricular zone in controls (Fig. 3b, c), whereas EphA7-electroporated migrating neurons had distinct segregation from EYFP[−] neurons in the intermediate zone, forming EYFP⁺ clusters (Fig. 3b, c). Live imaging observation confirmed that this segregation is mediated by tangential movement of neurons during the multipolar stage (Supplementary Movies 1 and 2).



d, Discontinuity of the band of NeuroD-labelled migrating neurons in the TKO cortex (arrowheads) by DTA electroporation. Similar levels of apoptotic cell death were observed in wild-type and TKO cortices as shown with adjacent slices. Note that a few surviving EYFP⁺ neurons can demarcate the electroporated fields (between dotted lines). Fibrous labelling in the upper intermediate zone in the NeuroD staining of TKO is a nonspecific background. **e**, Interruption of the NeuroD-labelled domain (arrows) in the TKO cortex, which received DTA electroporation only in a very narrow region. **f**, The thickness of Cutl1⁺ cortical layers at the DTA electroporated domains (brackets) was decreased in TKO. Relative thickness of Cutl1⁺ layers at the electroporated domains against the control hemisphere was significantly decreased in TKO. Data are mean \pm s.e.m. ($n = 20$ sections from five brains per genotype), $P = 0.0006$ (Student's t -test). Scale bars, 100 μ m (a), 50 μ m (b, c) and 200 μ m (d–f).

These results indicate that neurons segregate from each other owing to different levels of EphA and ephrin-A signalling that they receive during the multipolar phase of migration. Columnar grouping of EphA7-overexpressing neurons of multiclonal origins, through extensive tangential movement, has the opposite phenotype to that observed in TKO, in which tangential dispersion of cortical neurons is impaired and columnar arrays of clonally related neurons are formed (Fig. 2a, b). Taken together, our analyses indicate that EphA and ephrin-A signalling regulates lateral neuronal dispersion and intermingling during the multipolar stage of radial migration, and that this mechanism is required to generate cortical columns with appropriate cellular components (also see Supplementary Discussion). It is well established that progenitor cells have heterogeneous molecular properties, proliferative capacity and lineage commitment^{27–29}. We found that in wild-type, but not in TKO, an adjustment of neuron numbers, and an appropriate intermixing of neuronal subtypes can be achieved by lateral dispersion/intermixing of migrating neurons (Supplementary Fig. 10). These adaptive processes may be essential to achieve the modular function of cortical columns^{1,8,12}. Furthermore, given the preferential development of specific microcircuits within radially arrayed ontogenetic clones⁹, appropriate radial and tangential distribution of clonally related neurons may be essential for proper synaptic development and function of the radial units. The new role of EphA and ephrin-A signalling in assuring the homogeneous and continuous aspects of the neocortex demonstrated here contrasts with its previously known roles in restriction of cell movement and intermingling during formation of boundaries in other systems^{13,15,16}.

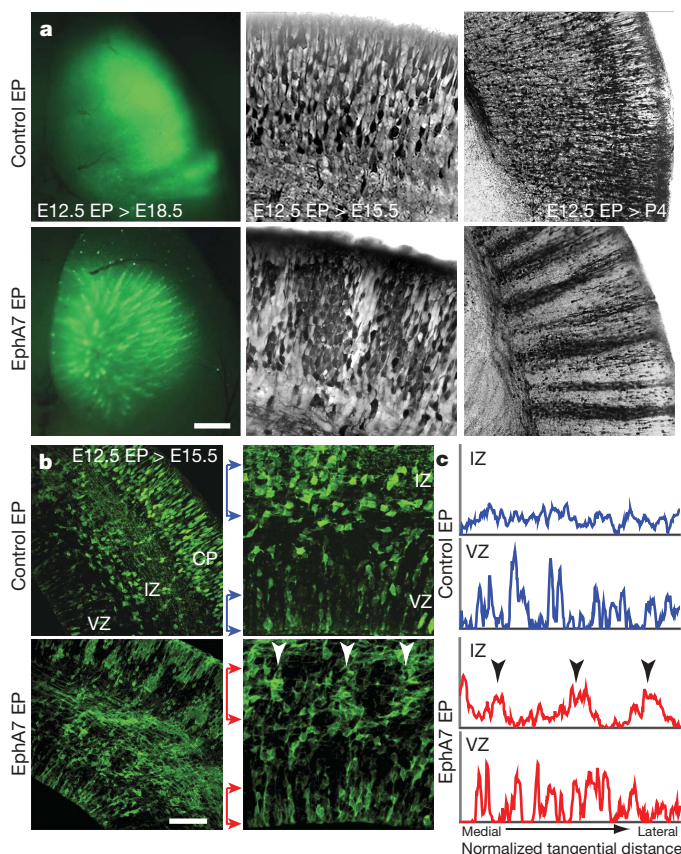


Figure 3 | EphA overexpression leads to tangential sorting of cortical neurons. **a**, Distribution of EYFP-labelled neurons (green or black) during development in the cortex in which control or EphA7 expression plasmid was delivered using electroporation (EP) with EYFP plasmid. Scale bars, 600 μm (for E18.8), 50 μm (for E15.5) and 200 μm (for P4). **b**, E15.5 cortex electroporated with control or EphA7 expression plasmid at E12.5. Right panels are higher magnification views around the ventricular-intermediate zone, showing multipolar shaped migrating neurons. EYFP⁺ EphA7-electroporated neurons start to segregate from EYFP⁻ neurons in the intermediate zone (arrowheads). Scale bars, 100 μm (left panels) and 50 μm (right panels). **c**, Fluorescence grey values were quantified from the images in **b** from left to right (medial to lateral) in the ventricular zone and the intermediate zone as indicated by blue and red arrows in **b**. Plot profiles indicate the clustering of labelled EphA7-electroporated neurons in the intermediate zone (arrowheads in **b** and **c**).

Although inappropriate neuronal positioning and abnormal columnar organization have been reported in post-mortem analysis of cortical tissue from subjects with neuropsychiatric disorders^{8,10,30}, the molecular mechanisms of these abnormalities remain unknown. Our finding provides a new perspective for investigating typical and atypical development that underlies higher order information processing in normal and pathophysiological states.

METHODS SUMMARY

Generation/genotyping of *Efna* TKO was performed as described¹⁸. Wild-type C57/Bl6 and CD-1 mice were used as controls. Histology, immunohistochemistry and *in situ* hybridization were performed on cryosections or vibratome sections as described previously²¹. Radial glial cells were labelled by injecting saturated DiI (1,1'-diiododecyl-3,3,3',3'-tetramethylindocarbocyanine perchlorate, Invitrogen) solution into the lateral ventricle of fixed P0 brains. CTPN and CPN were retrogradely labelled by injection of FluoroGold (Invitrogen) into the superior colliculus, and red fluorescent microspheres (FluoSperes, Invitrogen) into the contralateral parietal associative/visual cortex, respectively. GFP retrovirus (a gift from F. H. Gage) was injected *in utero* into the lateral ventricle of the embryo at E12.5. The distance between migrating neurons and their parental radial glia at E14.5 was quantified in Z-stacked confocal images using LSM Image Browser. *In utero* electroporation was performed at E12.5–E15.5. Each EphA and ephrin-A

expression plasmid was injected with pCAG-EYFP or pCAGGS-DsRed2. DTA expression plasmid was injected with pCAG-EYFP. The cortical area of each hemisphere was quantified using ImageJ software. Total numbers of cortical cells were estimated using the Isotropic Fractionator method. The thickness of Cutl1⁺ cortical layers was measured using ImageJ on coronal sections at the centre of electroporated domains and the corresponding region in the contralateral hemisphere, and the ratios were calculated. Pregnant mice (E13.5) were labelled with BrdU for 30 min and 24 h, and the labelling index (proportion of BrdU⁺ cells among Ki67⁺ cells) and cell-cycle exit index (proportion of Ki67⁻ cells among BrdU⁺ cells) were calculated. Apoptotic cell death was analysed by counting TUNEL⁺ cells in 10,000 μm^2 in the neocortex. For quantification of the pixel intensity along the tangential axis of the neocortex, the regions were delimited from medial to lateral in the neocortex, straightened with ImageJ, and the plot profiles were made.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 June; accepted 27 July 2009.

Published online 16 September 2009.

- Mountcastle, V. B. The columnar organization of the neocortex. *Brain* **120**, 701–722 (1997).
- Szentágothai, J. The Ferrier Lecture, 1977. The neuron network of the cerebral cortex: a functional interpretation. *Proc. R. Soc. Lond. B* **201**, 219–248 (1978).
- Goldman-Rakic, P. S. in *Handbook of Physiology, the Nervous System, Higher Functions of the Brain* (ed. Plum, F.) 373–417 (Am. Physiol. Soc., 1987).
- Rakic, P. Mode of cell migration to the superficial layers of fetal monkey neocortex. *J. Comp. Neurol.* **145**, 61–84 (1972).
- Rakic, P. Specification of cerebral cortical areas. *Science* **241**, 170–176 (1988).
- Rakic, P., Stensaas, L. J., Sayre, E. P. & Sidman, R. L. Computer-aided three-dimensional reconstruction and quantitative analysis of cells from serial electronmicroscopic montages of fetal monkey brain. *Nature* **250**, 31–34 (1974).
- Tan, S. S. & Breen, S. Radial mosaicism and tangential cell dispersion both contribute to mouse neocortical development. *Nature* **362**, 638–640 (1993).
- Buxhoeveden, D. P. & Casanova, M. F. The minicolumn hypothesis in neuroscience. *Brain* **125**, 935–951 (2002).
- Yu, Y. C., Bultje, R. S., Wang, X. & Shi, S. H. Specific synapses develop preferentially among sister excitatory neurons in the neocortex. *Nature* **458**, 501–504 (2009).
- Casanova, M. F. & Tillquist, C. R. Encephalization, emergent properties, and psychiatry: a minicolumnar perspective. *Neuroscientist* **14**, 101–118 (2008).
- Jones, E. G. Microcolumns in the cerebral cortex. *Proc. Natl Acad. Sci. USA* **97**, 5019–5021 (2000).
- de Kock, C. P. J., Bruno, R. M., Spors, H. & Sakmann, B. Layer- and cell-type-specific suprathreshold stimulus representation in rat primary somatosensory cortex. *J. Physiol. (Lond.)* **581**, 139–154 (2007).
- Xu, Q., Mellitzer, G., Robinson, V. & Wilkinson, D. G. *In vivo* cell sorting in complementary segmental domains mediated by Eph receptors and ephrins. *Nature* **399**, 267–271 (1999).
- Battle, E. *et al.* β -catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. *Cell* **111**, 251–263 (2002).
- Passante, L. *et al.* Temporal regulation of ephrin/Eph signalling is required for the spatial patterning of the mammalian striatum. *Development* **135**, 3281–3290 (2008).
- Mellitzer, G., Xu, Q. & Wilkinson, D. G. Eph receptors and ephrins restrict cell intermingling and communication. *Nature* **400**, 77–81 (1999).
- Gamble, J. A. *et al.* Disruption of ephrin signaling associates with disordered axophilic migration of the gonadotropin-releasing hormone neurons. *J. Neurosci.* **25**, 3142–3150 (2005).
- Pfeiffenberger, C. *et al.* Ephrin-As and neural activity are required for eye-specific patterning during retinogeniculate mapping. *Nature Neurosci.* **8**, 1022–1027 (2005).
- Cang, J. *et al.* Ephrin-As guide the formation of functional maps in the visual cortex. *Neuron* **48**, 577–589 (2005).
- Vanderhaeghen, P. *et al.* A mapping label required for normal scale of body representation in the cortex. *Nature Neurosci.* **3**, 358–365 (2000).
- Torii, M. & Levitt, P. Dissociation of corticothalamic and thalamocortical axon targeting by an EphA7-mediated mechanism. *Neuron* **48**, 563–575 (2005).
- Depaepe, V. *et al.* Ephrin signalling controls brain size by regulating apoptosis of neural progenitors. *Nature* **435**, 1244–1250 (2005).
- Tabata, H. & Nakajima, K. Multipolar migration: the third mode of radial neuronal migration in the developing cerebral cortex. *J. Neurosci.* **23**, 9996–10001 (2003).
- Maxwell, I. H., Maxwell, F. & Glode, L. M. Regulated expression of a diphtheria toxin A-chain gene transfected into human cells: possible strategy for inducing cancer cell suicide. *Cancer Res.* **46**, 4660–4664 (1986).
- Ge, W. *et al.* Coupling of cell migration with neurogenesis by proneural bHLH factors. *Proc. Natl Acad. Sci. USA* **103**, 1319–1324 (2006).
- Pasquale, E. B. Eph-ephrin bidirectional signaling in physiology and disease. *Cell* **133**, 38–52 (2008).

27. Soriano, E., Dumesnil, N., Auladell, C., Cohen-Tannoudji, M. & Sotelo, C. Molecular heterogeneity of progenitors and radial migration in the developing cerebral cortex revealed by transgene expression. *Proc. Natl Acad. Sci. USA* **92**, 11676–11680 (1995).
28. Gal, J. S. *et al.* Molecular and morphological heterogeneity of neural precursors in the mouse neocortical proliferative zones. *J. Neurosci.* **26**, 1045–1056 (2006).
29. Mizutani, K., Yoon, K., Dang, L., Tokunaga, A. & Gaiano, N. Differential Notch signalling distinguishes neural stem cells from intermediate progenitors. *Nature* **449**, 351–355 (2007).
30. Gleeson, J. G. & Walsh, C. A. Differential neuronal migration disorders: from genetic diseases to developmental mechanisms. *Trends Neurosci.* **23**, 352–359 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to T. Cutforth, D. A. Feldheim, J. G. Flanagan, J. Frisen, F. H. Gage, N. Y. Ip, K. Kohno, L. F. Kromer, C. Redies, J. L. R. Rubenstein and T. Saito for providing materials. We also thank M. R. Sarkisian and A. Bonnin for helpful comments, and M. Pappy, J. Bao, C. Anderson and S. Ellis for technical assistance. This work was supported by the NARSAD Young Investigator Award (M.T.), the Kavli Institute for Neuroscience at Yale (P.R.) and the National Institutes of Health (P.L. and P.R.)

Author Contributions M.T. initiated the project, conducted experiments, analysed the data, and wrote the manuscript. K.H-T. conducted experiments, analysed the data, and helped to write the manuscript. P.L. and P.R. contributed to the interpretation of results and writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.R. (pasko.rakic@yale.edu) and M.T. (masaaki.torii@yale.edu).

METHODS

Animals. All experiments using animals were in accordance with the protocols approved by Yale University Institutional Animal Care and Use Committee. The *Epha2/a3/a5* triple-knockout mouse was generated and genotyped as previously described^{18,31–33}. Wild-type mice were from the C57/Bl6 strain (Jackson). Wild-type CD-1 mice (Charles River) were also used in EphA and ephrin-A overexpression analysis. The day on which a vaginal plug was observed was designated as embryonic day 0.5.

In situ hybridization. *In situ* hybridization on cryosections or whole-mount was performed as described^{34,35}. The probes for *Epha5*, *Epha7*, *EfnA5* and *Cdh8* are described previously²¹. The probes for *Id2*, *Lmo3* and *Lmo4* are as described³⁶ (gifts from J. L. R. Rubenstein). The probes for *Epha4* (1,014-base-pair (bp) mouse partial clone), *EfnA2* (mouse full-length clone) and *EfnA3* (~200-bp rat partial clone) were gifts from L. F. Kromer.

Histology and immunohistochemistry. Brains were fixed with 4% paraformaldehyde (PFA) in PBS overnight, and 60–70- μ m coronal vibratome slices, or 20- μ m coronal cryosections, were collected. Immunohistochemistry was performed as described previously²¹. The following primary antibodies were used: polyclonal anti-GFP (also recognize EYFP; 1:3,000, Invitrogen; 1:250; Abcam), anti-DsRed (1:5,000; BD Biosciences), anti-Cutl1 (1:500; Santa Cruz Biotechnology), anti-Sox5 (1:5,000; Aviva Systems Biology), anti-Lmo4 (1:4,000; Abcam), anti-Laminin (1:1,000; Sigma-Aldrich), anti-NeuroD (1:300; Santa Cruz Biotechnology), monoclonal anti-CTIP2 (1:700; Abcam), anti-Ki67 (1:100; Neomarker), anti-BrdU (1:100 Beckton and Dickinson), anti-RC2 (1:5; DSHB), anti-Reelin (1:1,000; Millipore) and anti-MAP2 (1:200; Sigma-Aldrich) antibodies. Sections were nuclear counterstained with TO-PRO3 or 4'-6-diamidino-2-phenylindole (DAPI, Invitrogen) when necessary. Cresyl violet staining was performed using a standard protocol. All images were captured using a confocal LSM 510 NLO system or an Axioplan2 microscope (Carl Zeiss) equipped with epifluorescence.

Ligand binding and receptor binding histochemistry. Ligand and receptor binding histochemistry were performed as described²¹ using recombinant human EFNA5-Fc and EPHA7-Fc chimaeric protein (R&D Systems), respectively.

DiI, Fluorogold and fluorescent microsphere labelling. To label radial glial cells and their endfeet, saturated DiI solution (in ethanol) was injected into the lateral ventricle of P0 wild-type and TKO brains fixed with 4% PFA. After 48 h of incubation at 37 °C, coronal vibratome slices were made at 100- μ m thickness and photographed. CTPN and CPN in visual cortex were retrogradely labelled by pressure-injection of Fluorogold (Invitrogen) into the superior colliculus and red fluorescent microspheres (FluoSperes, Invitrogen) into the contralateral parietal associative/visual cortex, respectively, of P2 pups (to minimize the effect of activity-dependent reorganization). Coronal cryosections were made at P4 and images were captured using an Axioplan2 microscope (Carl Zeiss).

In utero retrovirus injection analysis. GFP retrovirus (a gift from F. H. Gage) was produced and injected *in utero* into E12.5 embryos as previously described³⁷. Brains were collected at E14.5 or P0, fixed in 4% PFA overnight, and thick vibratome sections (70 μ m) were made for GFP immunohistochemistry. To quantify the distance between migrating neurons and their parental radial glia, Z-stack images for each slice were taken using a confocal LSM 510 NLO system (Carl Zeiss), and the distance was measured using LSM Image Browser (Carl Zeiss). The investigator was blinded to the genotype for analysis.

Expression plasmids for electroporation. EphA7 expression plasmid (pCAGGS-EphA7-IRES-Venus), control plasmid (pCAGGS-IRES-EGFP) and pCAGGS-DsRed2 were described previously²¹. pCAG-EYFP and pCAG-DTA were provided by T. Saito³⁸ and K. Kohno³⁹, respectively. Full-length complementary DNA of mouse *Epha4* (a gift from N. Y. Ip), *EfnA2* and *EfnA5* (gifts from L. F. Kromer) were inserted into pCAGGS-IRES-EGFP as their expression plasmids. A truncated form of EphA4 (EphA4-T) lacking its intracellular domain (at Val 597) was designed as previously described⁴⁰, and a truncated form of EphA7 (EphA7-T) was similarly designed (truncated at Ile 609). cDNA fragments were amplified by PCR using full-length expression plasmids as templates, and the following primers: EphA4-T 5'-AGGAGCAGCGTTGGCACC-3' (forward) and 5'-CACCTAAGTTCTAACACCTTGATT-3' (reverse), EphA7-T 5'-CCATGGT TGTTCAACTCGGTA-3' (forward) and 5'-AATCTAGGTTTGGTGCCTG GA-3' (reverse), and the fragments were inserted into pCAGGS-IRES-EGFP.

In utero electroporation. *In utero* electroporation was performed as previously described³⁸ at E12.5–E15.5. Each EphA and ephrin-A expression plasmids (2 mg ml⁻¹) were injected with pCAG-EYFP (0.5 mg ml⁻¹) or pCAGGS-DsRed2 (1 mg ml⁻¹). In some cases, BrdU (50 mg kg⁻¹ body weight) was intraperitoneally injected immediately or 24 h after surgery for cell proliferation and cell-cycle exit analyses as described later. The DTA expression plasmid (1 mg ml⁻¹) was injected with pCAG-EYFP (2 mg ml⁻¹). We observed a massive increase in TUNEL⁺ cells in the electroporated domains of both wild-type and

TKO within 24 h ($n = 3$ and 2, respectively, data not shown). This striking increase of dying cells was restricted to the ventricular and intermediate zones, but was not present in the cortical plate at either E14.5 (data not shown) or E15.5 (Fig. 2d, $n = 6$ (wild-type) and 4 (TKO)). However, the increase in cell death was no longer evident by P0 (data not shown). Whole brain images were captured using Stemi SV 11 Apo stereomicroscope (Carl Zeiss) equipped with fluorescent source.

Organotypic slices and time-lapse imaging. Cortical slices were prepared as previously described⁴¹ from the E15.5 embryos electroporated at E13.5 ($n = 6$ each for control- and EphA7-electroporation). The slices were transferred to a RC25 recording chamber (Warner Instruments), maintained at 36 °C and perfused with fresh culture media. Time-lapse imaging of EYFP-labelled cells around the ventricular–intermediate zones was performed following the previously described procedure⁴². Time series were collected every 15 min for ~15 h per slice through a $\times 20$ objective on a confocal LSM 510 NLO system fitted with a motorized stage, and edited with Adobe Premier Pro.

Quantitative analyses of cortical size, Cutl1⁺ layer thickness, cell proliferation, cell-cycle exit and apoptotic cell death. The investigator was blinded to the genotype or electroporated gene. The cortical area of each hemisphere was quantified using ImageJ from images of the whole brain taken as dorsal view using Stemi SV 11 Apo stereomicroscope (Carl Zeiss), and averaged for wild-type and TKO. Total brain mass was weighed after fixation in 4% PFA overnight, and averaged for each genotype. The thickness of Cutl1⁺ cortical layers was measured using ImageJ on coronal sections at the centre of electroporated domains, and the relative thickness was calculated against the corresponding region in the non-electroporated hemisphere, and averaged for each genotype. A few sections that showed an apparent TKO phenotype of abnormal cortical thickness around the domain of measurement on either hemisphere were excluded from quantification. For proliferation and cell-cycle exit analyses, pregnant mice were injected intraperitoneally with BrdU (50 mg kg⁻¹ body weight) at E13.5, and embryos were collected after 30 min and 24 h, respectively. Brain sections were immunostained for Ki67 and BrdU, and the BrdU-labelling index (proportion of BrdU⁺ cells among Ki67⁺ cells) and cell-cycle-exit index (proportion of Ki67⁻ cells among BrdU⁺ cells) were calculated from the counted cell numbers as previously described⁴³. For quantification of apoptotic cell death, TUNEL assay was performed using ApopTag *in situ* apoptosis detection kit (Millipore) following the manufacturer's protocol, and TUNEL⁺ cells in 10,000 μ m² was calculated from counted cell numbers throughout the neocortex.

Quantification of the fluorescent pixel intensity along the tangential axis of the neocortex. The fluorescent pixel intensities of EPHA7- and EFNA5-Fc binding histochemistry in the intermediate zone (Supplementary Fig. 1), CTIP2 immunohistochemistry in the layer 5 (Fig. 1), and EYFP⁺ electroporated cells in the ventricular and intermediate zone (Fig. 3), along the medial-lateral axis of the neocortex, were quantified using a previously described method⁴⁴ with modification. Regions for quantification were delimited from medial to lateral in the neocortex, straightened with ImageJ software ('straighten' plugin), and the plot profiles of the fluorescent intensity was made as grey values (y) against the tangential distance (x). To present the distribution of EPHA-Fc and EFNA-Fc binding and EYFP⁺ electroporated cells, the pixel intensity was normalized and expressed as a percentage of the total value measured inside the region of interest. Pixel intensity of the CTIP2 labelling was not normalized between the wild-type and TKO cortex for comparison of their total values. Tangential distance was subgrouped into ~30 bins for better presentation of EPHA-Fc and EFNA-Fc binding and CTIP2 labelling.

Isotropic fractionator. Because of the highly anisotropic cerebral cortex of TKO, the total numbers of cells in a cortical hemisphere were estimated using the isotropic fractionator method as described⁴⁵ instead of using the stereological method. In brief, a cortical hemisphere was dissected from each wild-type and TKO brain at P0 and a suspension of nuclei was obtained by mechanical dissociation in 40 mM sodium citrate and 1% Triton X-100 using a tissue homogenizer. The homogenate was collected by centrifugation and suspended in PBS containing 1% DAPI, and the nuclear density was determined using a haemocytometer under a fluorescence microscope to estimate the total cell number in the original tissue. For each genotype, six pups from three dams were used. The coefficients of variation were <0.1 for both wild-type and TKO, indicating the reliability of the method. The investigator was blinded to the genotype for analysis.

31. Frisén, J. *et al.* Ephrin-A5 (AL-1/RAGS) is essential for proper retinal axon guidance and topographic mapping in the mammalian visual system. *Neuron* 20, 235–243 (1998).

32. Feldheim, D. A. *et al.* Genetic analysis of ephrin-A2 and ephrin-A5 shows their requirement in multiple aspects of retinocollicular mapping. *Neuron* 25, 563–574 (2000).

33. Cutforth, T. *et al.* Axonal ephrin-As and odorant receptors: coordinate determination of the olfactory sensory map. *Cell* 114, 311–322 (2003).

34. Grove, E. A., Tole, S., Limon, J., Yip, L. & Ragsdale, C. W. The hem of the embryonic cerebral cortex is defined by the expression of multiple *Wnt* genes and is compromised in *Gli3*-deficient mice. *Development* **125**, 2315–2325 (1998).
35. Suzuki, S. C., Inoue, T., Kimura, Y., Tanaka, T. & Takeichi, M. Neuronal circuits are subdivided by differential expression of type-II classic cadherins in postnatal mouse brains. *Mol. Cell. Neurosci.* **9**, 433–447 (1997).
36. Cholfin, J. A. & Rubenstein, J. L. Patterning of frontal cortex subdivisions by Fgf17. *Proc. Natl Acad. Sci. USA* **104**, 7652–7657 (2007).
37. Sanada, K., Gupta, A. & Tsai, L. H. Disabled-1-regulated adhesion of migrating neurons to radial glial fiber contributes to neuronal positioning during early corticogenesis. *Neuron* **42**, 197–211 (2004).
38. Saito, T. & Nakatsuji, N. Efficient gene transfer into the embryonic mouse brain using *in vivo* electroporation. *Dev. Biol.* **240**, 237–246 (2001).
39. Kimura, Y., Saito, M., Kimata, Y. & Kohno, K. Transgenic mice expressing a fully nontoxic diphtheria toxin mutant, not CRM197 mutant, acquire immune tolerance against diphtheria toxin. *J. Biochem.* **142**, 105–112 (2007).
40. Walkenhorst, J. *et al.* The EphA4 receptor tyrosine kinase is necessary for the guidance of nasal retinal ganglion cell axons in vitro. *Mol. Cell. Neurosci.* **16**, 365–375 (2000).
41. Haydar, T. F., Bambrick, L. L., Krueger, B. K. & Rakic, P. Organotypic slice cultures for analysis of proliferation, cell death, and migration in the embryonic neocortex. *Brain Res. Brain Res. Protoc.* **4**, 425–437 (1999).
42. Ang, E. S. Jr, Haydar, T. F., Gluncic, V. & Rakic, P. Four-dimensional migratory coordinates of GABAergic interneurons in the developing mouse cortex. *J. Neurosci.* **23**, 5805–5815 (2003).
43. Chenn, A. & Walsh, C. A. Regulation of cerebral cortical size by control of cell cycle exit in neural precursors. *Science* **297**, 365–369 (2002).
44. Bonnin, A., Torii, M., Wang, L., Rakic, P. & Levitt, P. Serotonin modulates the response of embryonic thalamocortical axons to netrin-1. *Nature Neurosci.* **10**, 588–597 (2007).
45. Herculano-Houzel, S. & Lent, R. Isotropic fractionator: a simple, rapid method for the quantification of total cell and neuron numbers in the brain. *J. Neurosci.* **25**, 2518–2521 (2005).

Asymptomatic deer excrete infectious prions in faeces

Gültekin Tamgüney^{1,2}, Michael W. Miller³, Lisa L. Wolfe³, Tracey M. Sirochman³, David V. Glidden⁴, Christina Palmer¹, Azucena Lemus⁵, Stephen J. DeArmond^{1,5} & Stanley B. Prusiner^{1,2}

Infectious prion diseases¹—scrapie of sheep² and chronic wasting disease (CWD) of several species in the deer family^{3,4}—are transmitted naturally within affected host populations. Although several possible sources of contagion have been identified in excretions and secretions from symptomatic animals^{5–8}, the biological importance of these sources in sustaining epidemics remains unclear. Here we show that asymptomatic CWD-infected mule deer (*Odocoileus hemionus*) excrete CWD prions in their faeces long before they develop clinical signs of prion disease. Intracerebral inoculation of irradiated deer faeces into transgenic mice overexpressing cervid prion protein (PrP) revealed infectivity in 14 of 15 faecal samples collected from five deer at 7–11 months before the onset of neurological disease. Although prion concentrations in deer faeces were considerably lower than in brain tissue from the same deer collected at the end of the disease, the estimated total infectious dose excreted in faeces by an infected deer over the disease course may approximate the total contained in a brain. Prolonged faecal prion excretion by infected deer provides a plausible natural mechanism that might explain the high incidence and efficient horizontal transmission of CWD within deer herds^{3,4,9}, as well as prion transmission among other susceptible cervids.

Prions are transmissible, proteinaceous agents that cause fatal neurodegenerative diseases¹. In cervids, including deer (*Odocoileus* spp.), elk (*Cervus elaphus nelsoni*) and moose (*Alces alces shirasi*), prions cause CWD^{4,10}. The incidence of CWD can be remarkably high both in captive and wild herds^{3,4,11} and epidemiological data argue that efficient horizontal transmission drives epidemic dynamics^{9,11,12}. Although cervids can be infected orally^{13,14} and seem to be able to contract CWD from contaminated environments¹⁵, precisely how and when CWD prions are shed into the environment have not been described. Previous studies have identified CWD prions in saliva, blood, urine, antler velvet, and muscle, lymphoid and other tissues of symptomatic cervids with late-stage disease^{5–7,14,16}. These sources of CWD prions may contribute to the spread of CWD, but none parsimoniously explains natural CWD transmission both within and between species in the deer family. To fit observed patterns, a natural CWD transmission mechanism must be effected within biologically realistic limits of the carrier medium, cannot require ‘cannibalism’^{9,15} and should be indirect to explain both environmental persistence and spread among multiple host species^{4,10,12,15}. Because empirical data^{13–15} and modelling¹² suggested faecal excretion of prions throughout much of the disease course as potentially important to CWD transmission, we investigated whether prions are shed in faeces from mule deer during the course of CWD infection.

To obtain a standard based on which we could determine the infectivity of tissues and excrement containing CWD prions, we first

performed endpoint titrations of the Elk1 CWD isolate¹⁷ by intracerebral inoculation into transgenic mice overexpressing elk prion protein (Tg(ElkPrP) mice) using 10 tenfold dilutions of a 10% brain homogenate ranging from 10^{-1} to 10^{-10} (Supplementary Fig. 1a). Based on the method of Spearman and Kärber¹⁸, we determined the titre of this homogenate to be 7.4 ± 0.24 log infectious (half maximal infectious dose (ID_{50}) \pm s.e.m.) units (U) ml^{-1} ; using Cox regression analysis¹⁹, which estimates the ID_{50} based on Kaplan–Meier survival times and accounts for censored events, the titre was 7.5 log ID_{50} U ml^{-1} with a 95% confidence interval (CI) between 7.0 and 7.9. Next, we correlated the ID_{50} values with the respective incubation times for dilutions 10^{-1} to 10^{-6} (fewer than 50% of the mice developed disease for dilutions greater than 10^{-7}) (Supplementary Fig. 1b) as a basis for determining unknown titres using the incubation time assay²⁰. The dynamic range of the incubation time assay strongly depends on the combination of animal model and prion strain used. In our system, dilutions of the Elk1 inoculum equal to or less than 10^{-3} showed similar incubation times, whereas dilutions greater than 10^{-7} did not cause disease, suggesting the dynamic range of this assay to be between 10^{-3} and 10^{-7} dilutions with median incubation times between 140 and 400 days (Supplementary Fig. 1a).

We assessed prion excretion in faeces collected from five mule deer before and every 3–6 months after oral infection with CWD prions until the animals died or developed signs of CWD and were euthanized at 16–20 months (Supplementary Table 1). The presence of CWD prions in these mule deer was confirmed at about 250 days post-inoculation (d.p.i.) by positive tonsil and rectal mucosa biopsies (Supplementary Table 1)²¹. Tg(ElkPrP) mice are highly susceptible to CWD prions of deer and seem to pose no species barrier despite one amino-acid difference at codon 226: elk express glutamate and deer PrP harbours glutamine^{17,22}. Therefore, we used bioassays to estimate CWD titres in brain tissue from these five deer (Fig. 1) as well as other orally infected deer with longer incubation times (Supplementary Fig. 2a), by intracerebral inoculation into Tg(ElkPrP) mice. We observed median incubation times of 131–207 days and estimated prion doses between 2.9 and 5.7 log ID_{50} units (Supplementary Table 2); titre estimates based on Cox regression analysis were similar (Supplementary Table 2). We found a linear correlation between the incubation period in deer and prion titres accumulated in their brains (Supplementary Fig. 2b).

In addition to brain homogenates, we performed bioassays using irradiated faecal homogenates collected from infected mule deer by intracerebral inoculation into Tg(ElkPrP) mice. Irradiation was used to damage nucleic acids and inactivate bacteria and viruses with minimal effects on prion titres²³; irradiation of the Elk1 CWD isolate did not diminish its titre when assayed in Tg(ElkPrP) mice (data not shown). Faecal samples collected from deer before oral infection and

¹Institute for Neurodegenerative Diseases, ²Department of Neurology, University of California, San Francisco, California, 94143 USA. ³Colorado Division of Wildlife, Wildlife Research Center, Fort Collins, Colorado, 80526 USA. ⁴Department of Epidemiology and Biostatistics, ⁵Department of Pathology, University of California, San Francisco, California, 94143 USA.

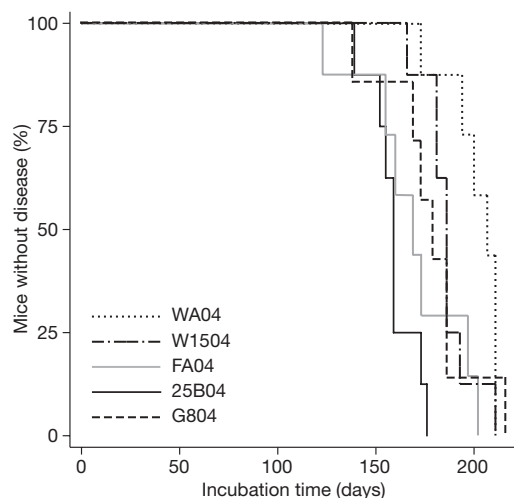


Figure 1 | Kaplan-Meier plots indicating incubation times in Tg(ElkPrP) mice after intracerebral inoculation with 1% (w/v) brain homogenate from mule deer (ID numbers indicated) that developed CWD in 16–20 months after oral infection with CWD prions.

3–4 months after infection did not transmit prion disease to Tg(ElkPrP) mice, whereas 14 of 15 faecal samples collected 4 months or later after inoculation transmitted disease to 37 of 126 Tg(ElkPrP) mice (29%; Fig. 2). Prion disease in mice with neurological symptoms was confirmed by detecting proteinase-K-resistant PrP^{Sc} in brain homogenates and in some cases by neuropathology (Fig. 3 and Supplementary Fig. 3). Neuropathological changes seen in symptomatic

Tg(ElkPrP) mice inoculated with faecal material collected from mule deer were similar to those in Tg(ElkPrP) mice inoculated with brain homogenates from the same CWD-positive mule deer (Fig. 4).

Transmission rates of CWD prions from some irradiated faecal samples to Tg(ElkPrP) mice were as high as 62% but most faecal homogenates caused disease in fewer than 50% of the inoculated mice. Compared with 30 μ l of a 10% brain homogenate from elk harbouring 5.9 log ID₅₀ units, we estimated that 30 μ l of 10% faecal homogenates contained between -1.6 and 0 log ID₅₀ units (that is, between 0.03 and 1 ID₅₀ units) (Supplementary Table 3 and Supplementary Note). Prion titres in the faeces from infected deer were variable for individual deer, but excretion of infectious prions was generally continuous until the deer became symptomatic, 16–20 months after oral infection. No trend was evident in infectivity levels in faeces over time beginning at 9 months after oral infection ($P = 0.99$). This observed prion-shedding pattern is consistent with the relatively early and rapid accumulation of prions in lymphoid tissue associated with the alimentary tract^{13,14}.

Faecal contamination of vegetation and soil provides a mechanism for transmitting a wide variety of parasitic bacteria, protozoa and helminths among herbivores²⁴. Faecal prion shedding also appears to be a plausible explanation for the efficient transmission of CWD. Based on our titration results, an infected deer may shed nearly as many prions in faeces over the disease course as accumulate in its brain in terminal disease. Assuming a constant infectious dose of 0 log ID₅₀ units in 3 μ g of faeces (equivalent to 5.5 log ID₅₀ units per gram of wet faeces), the cumulative total dose of prions shed during a 10-month period based on 780 g wet weight of faeces produced per day²⁵ is 10.9 log ID₅₀ units, which is similar to 10.8–12.3 log ID₅₀ units found in brains of terminally sick mule deer, assuming an average brain

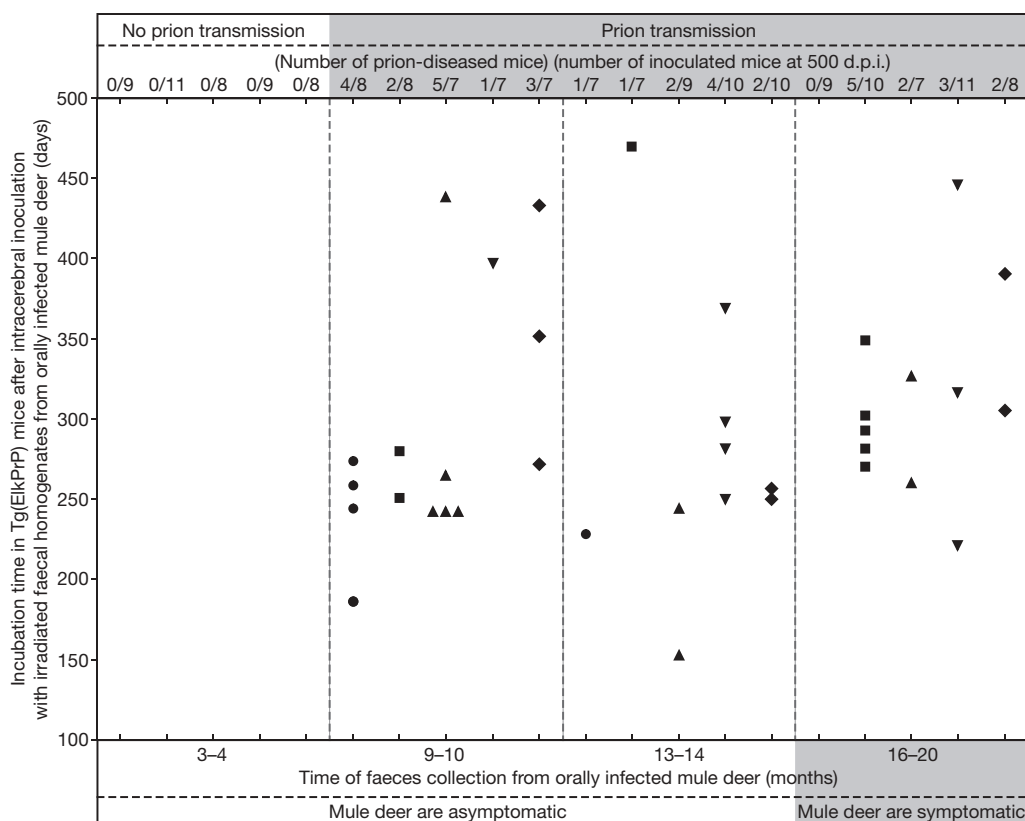


Figure 2 | Bioassays using irradiated faecal homogenates collected from infected mule deer by intracerebral inoculation into Tg(ElkPrP) mice.

Faeces from five mule deer (WA04, circles; W1504, squares; FA04, triangles; 25B04, inverted triangles; G804, diamonds) were sampled before and at four time points after oral infection; the last collection was taken when the deer developed signs of CWD. Irradiated faecal homogenates were inoculated

intracerebrally in Tg(ElkPrP) mice, some of which developed prion disease between 153 and 470 d.p.i. when inoculated with faecal preparations collected from deer 9 months or later after their oral infection. Faeces collected from deer before (data not shown) or at 3–4 months after oral infection did not transmit disease to Tg(ElkPrP) mice.

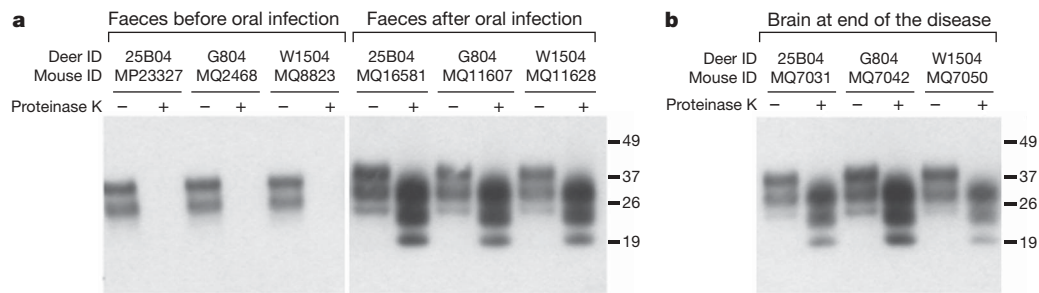


Figure 3 | Western blots of brain homogenates of Tg(ElkPrP) mice. Mice were inoculated intracerebrally with faeces (**a**) or brain homogenates (**b**) from mule deer. **a**, Mice inoculated with mule deer faeces collected before the deer were orally infected with CWD prions were killed at 481 d.p.i. and showed no proteinase-K-resistant PrP^{Sc} in their brains. In contrast, some mice inoculated with faeces from CWD-infected mule deer collected at

4 months or later after oral infection had PrP^{Sc} in their brains. **b**, Inoculation with brain homogenates from CWD-infected deer resulted in prion disease and PrP^{Sc} in the brains of ill Tg(ElkPrP) mice. Samples were undigested (–) or digested with proteinase K (+). Molecular masses of protein standards are shown in kilodaltons.

weight of 200 g. Although prion titres in faeces were relatively low, exposing deer to multiple oral doses in faeces-contaminated environments could increase their overall probability of infection²⁶. Moreover, both the persistence and infectivity of prions shed in faeces may be enhanced by interaction with clay soil microparticles²⁷, and mule deer may consume 8–30 g of soil daily depending on season²⁸, thereby facilitating infection by faecal prions. The foregoing mechanism is consistent with observed conditions under which captive

mule deer have shown remarkably high rates of prion infection^{3,4,9,12}, and explains how CWD could effectively transmit among mule deer in the wild.

Our findings show that CWD prions can be shed into the environment in faeces from symptomatic and, perhaps more importantly, from asymptomatic deer. These data support the faecal–oral route as a likely natural mechanism for the transmission of CWD prions among deer and other susceptible cervid species, and possibly for scrapie

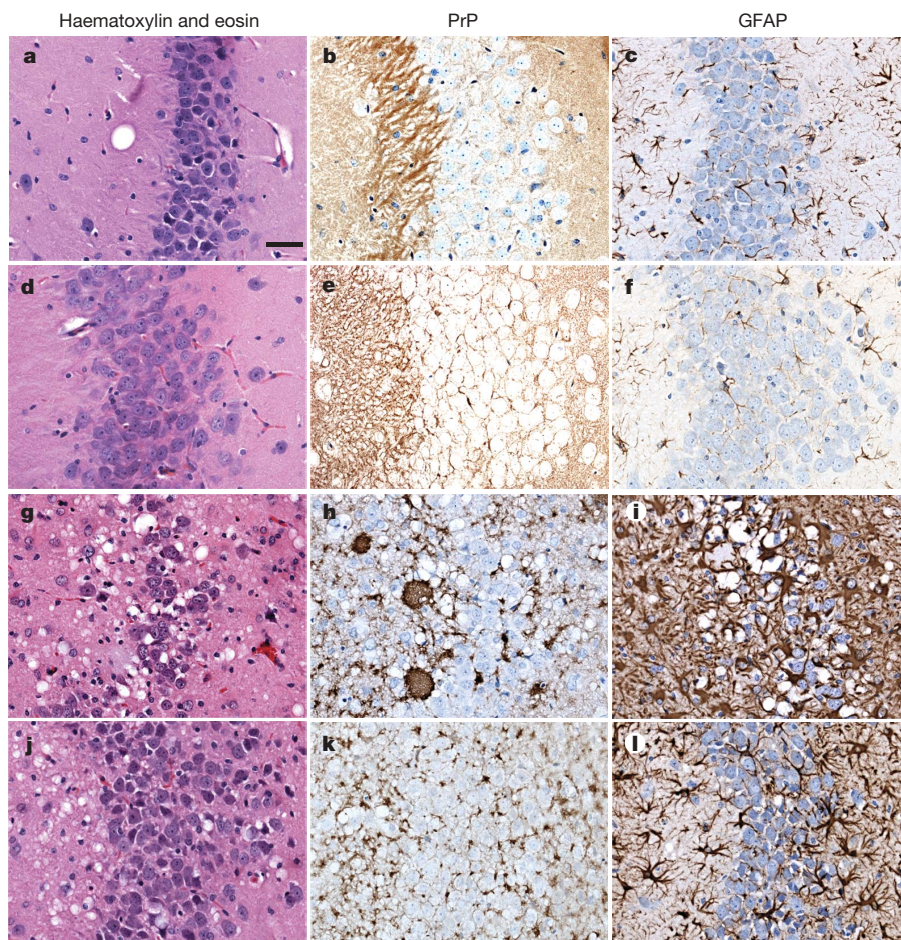


Figure 4 | Neuropathology of brain sections from Tg(ElkPrP) mice. Mice were inoculated with faecal (**d–i**) or brain (**j–l**) homogenates of mule deer. Sections show vacuolation (left), PrP^{Sc} deposition (middle) and astrocytic gliosis (right). Uninoculated 658-day-old control mice showed no vacuolation (**a**), no PrP^{Sc} deposits (**b**) and mild, age-related gliosis (**c**). Mice inoculated with faeces from uninfected deer remained healthy without

vacuolation (**d**) or PrP^{Sc} deposits (**e**), and showed mild, age-related gliosis at 503 d.p.i. (**f**). Mice inoculated with faeces from infected deer developed neurological symptoms in 153–470 days, showed vacuolation (**g**), PrP^{Sc} deposits (**h**), and severe gliosis (**i**) similar to mice inoculated with brain homogenates of deer with CWD (**j–l**). Scale bar, 25 μm.

prions among sheep and goats²⁹. Prion shedding through much of the disease course would facilitate exposure of both conspecifics and susceptible sympatric species, as well as geographic spread as deer move between seasonal ranges. Prion contamination of forest, shrub–steppe and grassland habitats may be largely responsible for horizontal transmission of CWD among mule deer and perhaps other species.

METHODS SUMMARY

Oral infection of mule deer. At weaning, mule deer fawns were orally infected with approximately 1 g of non-specific, pooled, infectious brain material placed at the base of the tongue; based on previous analyses, the inoculum pool contained approximately 3 µg PrP^{Sc} per gram of brain tissue³⁰ and showed prion conversion *in vitro* and infectivity *in vivo*^{13,14,30}. All infected deer surviving longer than 250 d.p.i. showed evidence of PrP^{Sc} accumulation in tonsil and rectal mucosa biopsies²¹, indicating successful infection. All infected deer that survived longer than 490 d.p.i. showed clinical signs of CWD before death and evidence of prion infection on post-mortem examination.

Mice and inoculations. Production of Tg(ElkPrP^{+/+})12584, or Tg(ElkPrP) for simplicity, has been described previously. Tg(ElkPrP) mice do not express endogenous mouse PrP and homozygously express the *PRNP* allele encoding ElkPrP(M132) from the cosSHa.Tet cosmid vector. For intracerebral inoculation, we injected 30 µl of diluted sample using a 27-gauge, disposable hypodermic syringe into the right parietal lobe of weanling mice. Inoculated animals were examined daily for their clinical status and three times weekly for neurological dysfunction and scored for prion disease based on standard diagnostic criteria. Diseased animals were euthanized and their brains removed. One half of the brain was frozen and the other half immersion-fixed in 10% neutrally buffered formalin for biochemical and neuropathological analyses, respectively. All mouse studies were reviewed and approved by the UCSF Institutional Animal Care and Use Committee.

Additional methods for care of mule deer, faeces sampling and treatment, statistical analyses, western immunoblotting and histopathology are described in the Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 May; accepted 16 July 2009.

Published online 9 September 2009.

1. Prusiner, S. B. in *Fields Virology* (eds Knipe, D. M. et al.) 3059–3092 (Lippincott Williams & Wilkins, 2007).
2. Detwiler, L. A. & Baylis, M. The epidemiology of scrapie. *Rev. Sci. Tech.* **22**, 121–143 (2003).
3. Williams, E. S. & Young, S. Chronic wasting disease of captive mule deer: a spongiform encephalopathy. *J. Wildl. Dis.* **16**, 89–98 (1980).
4. Williams, E. S. Chronic wasting disease. *Vet. Pathol.* **42**, 530–549 (2005).
5. Mathiason, C. K. et al. Infectious prions in the saliva and blood of deer with chronic wasting disease. *Science* **314**, 133–136 (2006).
6. Angers, R. C. et al. Chronic wasting disease prions in elk antler velvet. *Emerg. Infect. Dis.* **15**, 696–703 (2009).
7. Haley, N. J., Seelig, D. M., Zabel, M. D., Telling, G. C. & Hoover, E. A. Detection of CWD prions in urine and saliva of deer by transgenic mouse bioassay. *PLoS ONE* **4**, e4848 (2009).
8. Konold, T., Moore, S. J., Bellworthy, S. J. & Simmons, H. A. Evidence of scrapie transmission via milk. *BMC Vet. Res.* **4**, 14 (2008).
9. Miller, M. W. & Williams, E. S. Prion disease: horizontal prion transmission in mule deer. *Nature* **425**, 35–36 (2003).
10. Baeten, L. A., Powers, B. E., Jewell, J. E., Spraker, T. R. & Miller, M. W. A natural case of chronic wasting disease in a free-ranging moose (*Alces alces shirasi*). *J. Wildl. Dis.* **43**, 309–314 (2007).
11. Miller, M. W. et al. Lions and prions and deer demise. *PLoS ONE* **3**, e4019 (2008).

12. Miller, M. W., Hobbs, N. T. & Tavenner, S. J. Dynamics of prion disease transmission in mule deer. *Ecol. Appl.* **16**, 2208–2214 (2006).
13. Sigurdson, C. J. et al. Oral transmission and early lymphoid tropism of chronic wasting disease PrP^{res} in mule deer fawns (*Odocoileus hemionus*). *J. Gen. Virol.* **80**, 2757–2764 (1999).
14. Fox, K. A., Jewell, J. E., Williams, E. S. & Miller, M. W. Patterns of PrP^{CWD} accumulation during the course of chronic wasting disease infection in orally inoculated mule deer (*Odocoileus hemionus*). *J. Gen. Virol.* **87**, 3451–3461 (2006).
15. Miller, M. W., Williams, E. S., Hobbs, N. T. & Wolfe, L. L. Environmental sources of prion transmission in mule deer. *Emerg. Infect. Dis.* **10**, 1003–1006 (2004).
16. Angers, R. C. et al. Prions in skeletal muscles of deer with chronic wasting disease. *Science* **311**, 1117 (2006).
17. Tamgüney, G. et al. Transmission of elk and deer prions to transgenic mice. *J. Virol.* **80**, 9104–9114 (2006).
18. Dougherty, R. in *Techniques in Experimental Virology* (ed. Harris, R. J. C.) 169–224 (Academic, 1964).
19. Cox, D. R. Regression models and life-tables. *J. R. Stat. Soc. A* **34**, 187–220 (1972).
20. Prusiner, S. B. et al. Measurement of the scrapie agent using an incubation time interval assay. *Ann. Neurol.* **11**, 353–358 (1982).
21. Wolfe, L. L. et al. PrP^{CWD} in rectal lymphoid tissue of deer (*Odocoileus* spp.). *J. Gen. Virol.* **88**, 2078–2082 (2007).
22. Tamgüney, G. et al. Transmission of scrapie and sheep-passaged bovine spongiform encephalopathy prions to transgenic mice expressing elk prion protein. *J. Gen. Virol.* **90**, 1035–1047 (2009).
23. Miekka, S. I. et al. Inactivation of viral and prion pathogens by gamma-irradiation under conditions that maintain the integrity of human albumin. *Vox Sang.* **84**, 36–44 (2003).
24. Judge, J., Greig, A., Kyriazakis, I. & Hutchings, M. R. Ingestion of faeces by grazing herbivores—risk of inter-species disease transmission. *Agric. Ecosyst. Environ.* **107**, 267–274 (2005).
25. Arthur, W. J. Jr & Alldredge, A. W. Seasonal estimates of masses of mule deer fecal pellets and pellet groups. *J. Wildl. Mgmt* **44**, 750–752 (1980).
26. Diringer, H., Roehmel, J. & Beekes, M. Effect of repeated oral infection of hamsters with scrapie. *J. Gen. Virol.* **79**, 609–612 (1998).
27. Johnson, C. J., Pedersen, J. A., Chappell, R. J., McKenzie, D. & Aiken, J. M. Oral transmissibility of prion disease is enhanced by binding to soil particles. *PLoS Pathog.* **3**, e93 (2007).
28. Arthur, W. J. III & Alldredge, A. W. Soil ingestion by mule deer in northcentral Colorado. *J. Range Mgmt* **32**, 67–71 (1979).
29. Hadlow, W. J., Kennedy, R. C. & Race, R. E. Natural infection of Suffolk sheep with scrapie virus. *J. Infect. Dis.* **146**, 657–664 (1982).
30. Raymond, G. J. et al. Evidence of a molecular barrier limiting susceptibility of humans, cattle and sheep to chronic wasting disease. *EMBO J.* **19**, 4425–4430 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements G.T. was supported by a fellowship from the Larry L. Hillblom Foundation. This work was supported by the Colorado Division of Wildlife and grants from the US Department of Defense National Prion Research Program (NP020152), the National Institutes of Health (AG02132) and the Hillblom Foundation. We thank P. Lessard and the staff of the Hunters Point animal facility for support with the transgenic animal experiments, K. Giles for screening transgenic animals, A. Serban for antibodies, K. Pomeroy for technical assistance, J. Safar for discussions, H. Nguyen for editorial assistance and B. Williams for insights on CWD transmission.

Author Contributions G.T., M.W.M. and S.B.P. designed the transgenic mouse studies; G.T., M.W.M., L.L.W., T.M.S., C.P., A.L. and S.J.D. performed various aspects of the research on mule deer or transgenic mice; G.T., M.W.M., D.V.G., S.J.D. and S.B.P. analysed the data; G.T., M.W.M., S.J.D. and S.B.P. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.B.P. (E-mail: stanley@ind.ucsf.edu.)

METHODS

Housing and care of mule deer. Captive mule deer were held at the Colorado Division of Wildlife's Foothills Wildlife Research Facility as part of an ongoing study on prion shedding patterns in North American cervid species^{21,31}. Mule deer fawns were acquired and bottle-raised using canned evaporated bovine milk and established protocols. Deer were confined to 0.1-hectare (1,000-m²) biosecure paddocks throughout the study, except when held in metabolic cages for sample collections. Alfalfa hay, pelleted supplement (Ranch-Way deer diet and Mazuri 'browser' ration), mineralized salt blocks and water were provided *ad libitum* in all paddocks as per standard feeding and husbandry protocols.

Faeces sampling. Faecal samples were obtained from deer before oral infection, then approximately every 6 weeks on a rotating basis (Supplementary Table 1). During sample collection, deer were held in individual metabolic cages (1.5 m × 2.5 m) for two consecutive days; feed and water were provided *ad libitum*. During this period, 100 g of voided faeces were collected daily, composited and stored at -20 °C for later analysis. We selected samples of faeces representing relatively early, middle and late stages of the CWD course as well as terminal disease to assess the occurrence and relative timing of prion shedding in faeces.

Treatment of faeces from CWD-infected mule deer. Faecal samples collected from mule deer were kept frozen in plastic containers at -80 °C. To prepare a 14% (w/v) homogenate of faeces, 1 g of frozen faeces was thawed and homogenized by vortexing for 15 min in diluent containing 5% (w/v) bovine albumin Fraction V (ICN) in Mg²⁺- and Ca²⁺-free PBS, pH 7.4. Large insoluble particles were removed by filtering the homogenate through a 100-μm cell strainer (Becton Dickinson). After additional homogenization using a syringe with a 21-gauge needle, a 1-ml aliquot was transferred to a sterile, plastic, screw-cap tube that was irradiated by cesium-137 at a total dose of 7.0 kGy on dry ice, which has a negligible effect on prion infectivity but inactivates bacteria and viruses^{23,32}. Irradiated 14% faecal homogenates were tested on 5% sheep blood agar plates for residual microbial activity before they were intracerebrally inoculated into Tg(ElkPrP) mice.

Statistical analysis of prion infectivity in transgenic mice. Production of Tg(ElkPrP^{+/+})12584 has been described^{33,34}. Inoculated mice were examined daily for their clinical status, three times weekly for neurological dysfunction and scored for prion disease based on standard diagnostic criteria^{35,36}. The method of Spearman and Kärber¹⁸ was used to calculate the ID₅₀ value of a 10% brain homogenate from an elk diagnosed with CWD, and denoted the Elk1 isolate¹⁷. We used two approaches to calibrate titres in 10% brain and 14% faecal homogenates from mule deer to an equivalent titre of 10% brain homogenate of Elk1. The first approach was based on a nonlinear regression model for the mean times to illness for the animals. Such an approach is only applicable when all animals fall ill during the observation period. In the second, more flexible model, the statistical methods of survival analysis were used to compare onset times. Survival analysis methods³⁷ can handle observations that are 'censored', such as animals that die from competing, unrelated causes or that do not become ill during the study period. This method is advantageous when attempting to calibrate an inoculum that does not result in illness in all animals. The methods operate by considering $k = 1, \dots, K$ different experiments to be compared with the serial dilution series of Elk1. Those experiments are then paired with the serial Elk1 dilution data. For the i th comparison, the variable group _{k} is 1 for animals in the k th experiment, 0 for animals in the Elk1 serial dilution series, and dilution _{k} gives the dilution in the k th experiment relative to a 10% homogenate. The nonlinear regression model for the mean onset time in the k th series, denoted t_k , is $\log(t_k) = \alpha_0 + \alpha_1 \text{group}_k + \alpha_2 \text{dilution}_k$. The equivalent log titre for the k th experiment series relative to the ID₅₀ of the Elk1 isolate is then

$\alpha_1 \alpha_2 - \log_{10}(\text{ID}_{50} \text{ Elk1})$. For cases in which not all animals die, the mean illness time cannot be defined, but the hazard function, which describes the rate of the onset of disease, can be defined. We used the Cox proportional hazards model¹⁹ to represent the hazard as $h_k(t) = h_0(t) \exp(\beta_1 \text{group}_k + \beta_2 \text{dilution}_k)$. The equivalent log titre for the k th experiment series relative to the ID₅₀ of the Elk1 isolate is then $\beta_1/\beta_2 - \log_{10}(\text{ID}_{50} \text{ Elk1})$. All calculations were performed with Stata 10 (Stata Corp.); the annotated analysis code is shown in the Supplementary Note. We analysed data on incubation periods for affected-faeces-inoculated Tg(ElkPrP) mice by logistic regression³⁸ to look for trends in incubation periods that might suggest temporal trends in prion shedding in mule deer faeces.

Western blotting. For western immunoblotting analysis, 10% (w/v) brain homogenates were prepared in PBS by two 45-s runs at 6.0 m s⁻¹ with a FastPrep FP120 Cell Disrupter (Qbiogene). Samples of 5% brain homogenates were incubated with 20 μg ml⁻¹ of proteinase K (New England Biolabs) for 1 h at 37 °C. Proteinase K digestion was stopped with 2 mM phenylmethylsulphonyl fluoride and samples were centrifuged at 100,000g for 1 h at 4 °C. Pellets were resuspended in 10 mM Tris-HCl (pH 8.0), 0.15 M NaCl, 0.5% (w/v) NP-40, 0.5% (w/v) sodium deoxycholate. Equal volumes of 2 × SDS sample buffer were added to the samples before they were boiled for 5 min. For electrophoresis, 30 μl of undigested and proteinase-K-digested samples were loaded onto the gels³⁹. SDS gel electrophoresis and western blotting were performed using 12% Novex or 4–12% Tris-glycine gels and the iBlot dry blotting system (Invitrogen). PrP was detected with the human-mouse (HuM) recombinant antibody fragment (recFab) P covalently bound to horseradish peroxidase and developed with the enhanced chemiluminescent detection system (Amersham Biosciences)^{34,40}.

Histopathology. Tissue samples were either immediately frozen after removal or immersion-fixed in 10% buffered formalin and embedded in paraffin. To evaluate vacuolation, brain sections 8 μm thick were stained with haematoxylin and eosin. Reactive astrocytic gliosis was evaluated by immunostaining for glial fibrillary acidic protein using a rabbit antiserum (Dako). Detection of PrP^{Sc} from formalin-fixed tissue sections was performed with the HuM-P mouse PrP-specific monoclonal antibody after hydrolytic autoclaving, as previously described⁴¹.

31. Williams, E. S. *et al.* *Epidemiology of Chronic Wasting Disease: Detection of PrP^{res}, Shedding, and Environmental Contamination* (United States Department of Defense National Prion Research Program, 2003).
32. Bellinger-Kawahara, C. G., Kempner, E., Groth, D. F., Gabizon, R. & Prusiner, S. B. Scrapie prion liposomes and rods exhibit target sizes of 55,000 Da. *Virology* **164**, 537–541 (1988).
33. Büeler, H. *et al.* Normal development and behaviour of mice lacking the neuronal cell-surface PrP protein. *Nature* **356**, 577–582 (1992).
34. Tamgüney, G. *et al.* Genes contributing to prion pathogenesis. *J. Gen. Virol.* **89**, 1777–1788 (2008).
35. Carlson, G. A. *et al.* Genetics and polymorphism of the mouse prion gene complex: control of scrapie incubation time. *Mol. Cell. Biol.* **8**, 5528–5540 (1988).
36. Scott, M. *et al.* Propagation of prions with artificial properties in transgenic mice expressing chimeric PrP genes. *Cell* **73**, 979–988 (1993).
37. Cox, D. R. & Oakes, D. *Analysis of Survival Data* (Chapman & Hall and CRC, 1984).
38. Vittinghoff, E., Glidden, D. V., Shiboski, S. C. & McCulloch, C. E. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models* (Springer Science+Business Media, 2005).
39. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
40. Safar, J. G. *et al.* Measuring prions causing bovine spongiform encephalopathy or chronic wasting disease by immunoassays and transgenic mice. *Nature Biotechnol.* **20**, 1147–1150 (2002).
41. Muramoto, T. *et al.* Heritable disorder resembling neuronal storage disease in mice expressing prion protein with deletion of an α -helix. *Nature Med.* **3**, 750–755 (1997).

Fgf8 morphogen gradient forms by a source-sink mechanism with freely diffusing molecules

Shuizi Rachel Yu^{1,3*}, Markus Burkhardt^{2,3*}, Matthias Nowak^{1,3}, Jonas Ries^{2,3}, Zdeněk Petrášek^{2,3}, Steffen Scholpp^{1†}, Petra Schwillé^{2,3} & Michael Brand^{1,3}

It is widely accepted that tissue differentiation and morphogenesis in multicellular organisms are regulated by tightly controlled concentration gradients of morphogens^{1,2}. How exactly these gradients are formed, however, remains unclear^{3–12}. Here we show that Fgf8 morphogen gradients in living zebrafish embryos are established and maintained by two essential factors: fast, free diffusion of single molecules away from the source through extracellular space, and a sink function of the receiving cells, regulated by receptor-mediated endocytosis. Evidence is provided by directly examining single molecules of Fgf8 in living tissue by fluorescence correlation spectroscopy, quantifying their local mobility and concentration with high precision. By changing the degree of uptake of Fgf8 into its target cells, we are able to alter the shape of the Fgf8 gradient. Our results demonstrate that a freely diffusing morphogen can set up concentration gradients in a complex multicellular tissue by a simple source-sink mechanism.

Morphogens are signalling molecules that can induce distinct cellular responses in a concentration-dependent manner^{1,2}. Many examples have previously been described, including fibroblast growth factor 8 (Fgf8) in zebrafish embryonic development^{2,13–18}. How morphogens such as Fgf8 move through a tissue and set up concentration gradients remains debated. In the 1970s, Francis Crick proposed that simple diffusion and spatially uniform degradation is sufficient for setting up gradients³. This mechanism has been considered in various systems^{6–11} and it has been shown that a random walk model can accurately describe the DPP morphogen gradient in *Drosophila* wing discs¹⁰. Moreover, studies have suggested that morphogens can be sequestered from target tissue by receptor-mediated removal^{12,18}. However, a full biophysical description of morphogen gradient formation is still lacking. Meanwhile, alternative models including receptor-aided bucket-brigade mechanism⁴ and directed transport¹¹ have been proposed for morphogen spreading. Here we report a fluorescence correlation spectroscopy (FCS)-based assay for examining single-molecule dynamics of Fgf8 in living zebrafish embryos (Fig. 1a–c and Supplementary Information)^{19,20}. By directly measuring its mobility and concentration gradient, we determine the mechanism of Fgf8 morphogen gradient formation (Supplementary Fig. 1).

We first tagged Fgf8 with enhanced green fluorescent protein (Fgf8–EGFP; Fig. 2a). At equimolar concentrations, the fusion protein showed a similar capacity to induce target gene expression as untagged Fgf8 (Supplementary Fig. 2a, b). Fgf8–EGFP was readily taken up by target cells, and was recruited onto cell membranes by Fgf receptors (Fig. 2a–e)²¹. In contrast, a negative control of secreted EGFP (secEGFP) was not. Furthermore, we verified by western blot that the full-length fusion protein was the only EGFP-containing species in the extracellular space of *fgf8-egfp* messenger-RNA-injected embryos (Supplementary Fig. 2c).

To investigate how Fgf8 spreads into its target tissue, we took FCS measurements in the extracellular space of gastrulating zebrafish embryos that had a restricted source of the fluorescent protein (Figs 1 and 2b, c and Methods Summary). The FCS autocorrelation curves fit well to a three-dimensional diffusion model, but not with alternative models such as directional active transport only (Fig. 3a and Supplementary Information). Most of the molecules (91%) had a diffusion coefficient ($53 \mu\text{m}^2 \text{s}^{-1}$) that was on the same order of magnitude as that of secEGFP ($86 \mu\text{m}^2 \text{s}^{-1}$; Fig. 3a, b and Supplementary Fig. 3a, b). Fgf8–EGFP moved slightly slower than secEGFP, as is expected for a larger molecule of about double the molecular mass. For a receptor-mediated bucket-brigade mechanism⁴, we would expect most of the ligand to have a diffusion coefficient similar to that of the lateral mobility of the Fgf receptor in the cell membrane, which is about 100-fold smaller²² than the one we measured for Fgf8, thus arguing against this mechanism for Fgf8 spreading. The Fgf8–EGFP diffusion coefficient in the zebrafish embryos is two orders of magnitude larger than that reported for DPP ($0.10 \mu\text{m}^2 \text{s}^{-1}$) in *Drosophila* wing discs¹⁰. This supports the notion that the two morphogens have very different propagation mechanisms in their respective target tissues. DPP is transported by transcytosis involving repeated cycles of endocytosis and exocytosis through target cells^{5,10}. This is likely to occur much slower than the free diffusion of Fgf8–EGFP in the extracellular space.

In addition, we analysed recombinant Fgf8 protein labelled with the organic dye Cy5. The autocorrelation curve of Fgf8–Cy5 fitted

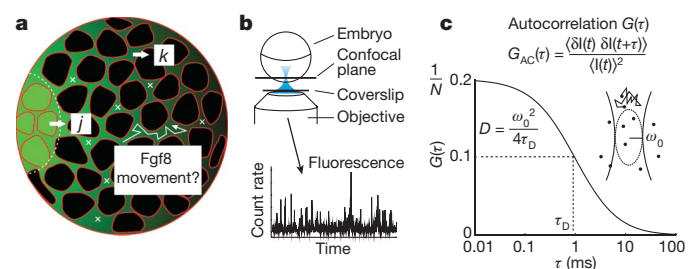


Figure 1 | FCS setup to investigate morphogen characteristics. **a**, Schematic section of a gastrulating zebrafish embryo. Morphogen is produced from a restricted source (green cells) with current j , spreads into target tissue (black cells) and is removed from the extracellular space with degradation rate k . Crosses indicate extracellular spaces where FCS measurements are taken. **b**, One-focus FCS setup. **c**, The autocorrelation (AC) function $G(\tau)$ has two key parameters: (1) the characteristic dwell time τ_D , which relates to the diffusion coefficient D ; and (2) the amplitude $G(0)$, which is inversely proportional to the number of fluorescent particles N in the detection volume (Supplementary Information).

¹Developmental Genetics, Biotechnology Center, TUD, ²Biophysics, Biotechnology Center, TUD, and ³Center for Regenerative Therapies, TUD, Tatzberg 47–49, 01307 Dresden, Germany. [†]Present address: Institute of Toxicology and Genetics, Forschungszentrum Karlsruhe, D-76021 Karlsruhe, Germany.

*These authors contributed equally to this work.

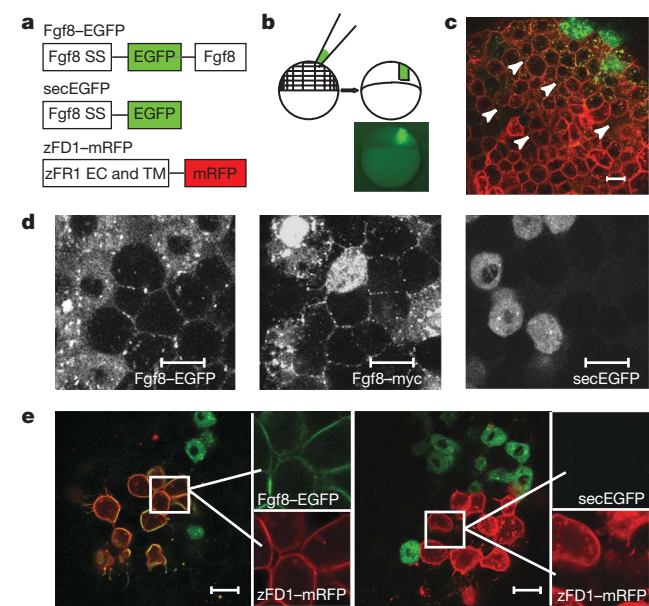


Figure 2 | Labelling strategy for Fgf8. **a**, Fgf8-EGFP, secEGFP and zFD1-mRFP (monomeric red fluorescent protein) constructs. SS, zebrafish Fgf8 signal peptide (1–22 amino acids); zFR1, zebrafish Fgf receptor 1. EC and TM denote zFR1 extracellular and transmembrane domains (1–390 amino acids), respectively. **b**, Blastomere injection and an injected embryo at sphere stage. **c**, Confocal image of a blastomere-injected embryo at 30% gastrula stage. Green: Fgf8-EGFP; red: mRFP-glycosyl phosphatidylinositol (GPI; a membrane marker); arrowheads denote extracellular spaces where FCS measurements are taken. **d**, In blastomere-injected embryos, Fgf8-EGFP and Fgf8-myc are taken up into vesicular structures in target cells, whereas secEGFP is not. **e**, Fgf8-EGFP accumulates on zFD1-overexpressing cells, whereas secEGFP does not. Green: Fgf8-EGFP or secEGFP; red: zFD1-mRFP. Scale bars, 20 μm .

well to a three-dimensional diffusion model with a diffusion coefficient of $91 \mu\text{m}^2 \text{s}^{-1}$ (Fig. 3b and Supplementary Fig. 3c). This shows that different fluorescent tags did not affect the mechanism of Fgf8 propagation. The similarity of the diffusion coefficients of Fgf8-EGFP, Fgf8-Cy5 and secEGFP to that of EGFP in buffer

($95 \mu\text{m}^2 \text{s}^{-1}$)²³ suggests that these molecules all move as single molecules by free Brownian diffusion through the extracellular space.

We have ruled out the directional active-transport-only model for Fgf8 tissue propagation (Fig. 3a). However, to examine whether Fgf8 movement has any active transport component, we used two-focus FCS. This is a directional-sensitive extension to the standard one-focus FCS^{24,25}. Here the fluorescence signals from two laterally shifted confocal volumes are cross-correlated to give a measure for the probability that a molecule that has been detected in one focal volume is subsequently detected in the other (Supplementary Fig. 4a, b and Supplementary Information). When the auto- and cross-correlation curves of Fgf8-EGFP were globally fit to a model including both diffusion and directed transport (Fig. 3c, Supplementary Equation 7 and Supplementary Information), we found that the forward and backward cross-correlation curves were identical within the experimental error and we obtained an average directional speed of $v = (2 \pm 4) \mu\text{m s}^{-1}$ (mean \pm s.d., $n = 10$). This shows that Fgf8 movement is non-directional and agrees with the fact that free diffusion is a more efficient mechanism than directed motion for a molecule to traverse complex multicellular tissue⁹.

We detected a small fraction of the extracellular Fgf8-EGFP (9%) that was moving significantly slower ($4 \mu\text{m}^2 \text{s}^{-1}$; Fig. 3b, Supplementary Fig. 3b and Supplementary Information) than the diffusing majority. This fraction was also present for Fgf8-Cy5 but not for secEGFP (Fig. 3b and Supplementary Fig. 3a, c), suggesting Fgf8-specific interactions with extracellular matrix components. A candidate for such interactions is heparan sulphate proteoglycan (HSPG), which has been suggested to interact with and to be required for the distribution and function of many signalling molecules^{26,27}. We injected heparinase I (HepI)²⁸ enzyme into the extracellular space of zebrafish embryos to cleave the sugar chains of HSPGs. Phenotypically, HepI treatment resulted in the expansion of endogenous expression domains of Fgf target genes in an Fgf-receptor-dependent manner and overall dorsalization similar to an Fgf gain-of-function situation (Supplementary Fig. 5a–k and Supplementary Information). This indicates that HSPGs are important for restricting the signalling range of Fgf proteins. FCS measurements showed a one-third reduction in the amount of the slow-moving Fgf8 in the HepI-treated embryos. On the other hand, the injection of heparan sulphate (the sugar chain of HSPG) led to significant increase in the slow-moving Fgf8 component (8.6% control, 5.6% HepI and 36%

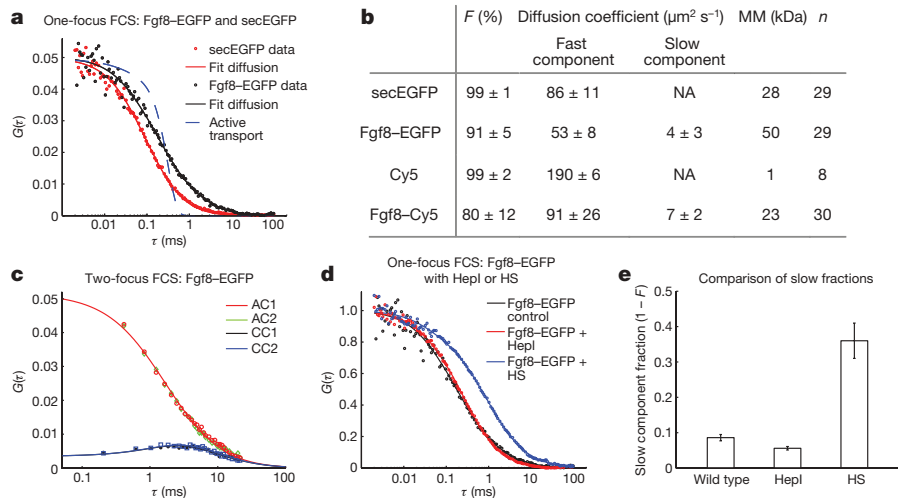


Figure 3 | Fgf8-EGFP diffuse as single molecules in the extracellular space. **a**, One-focus autocorrelation curves for secEGFP (red) and Fgf8-EGFP (black) fit with three-dimensional diffusion model; the Fgf8-EGFP autocorrelation curve is also fit with a directional transport-only model (blue). **b**, Fit results of autocorrelation curves measured in zebrafish embryos. F, fraction of the fast-diffusing component; MM, molecular mass (without signal peptide). Data are mean \pm s.d. of n measurements.

c, Two-focus autocorrelation (red and green) and cross-correlation (blue and black) data for Fgf8-EGFP fit with a diffusion-directed transport model. **d**, One-focus Fgf8-EGFP autocorrelation curves in the presence of HepI (red) or heparan sulphate (HS; blue). **e**, Effect of HepI and heparan sulphate on the fraction of slow-moving Fgf8 (Supplementary Information). Error bars in **e** denote s.e.m.

heparan sulphate; Fig. 3d, e and Supplementary Information). Together, we propose that most of the extracellular Fgf8 are freely diffusing and this allows the molecules to spread quickly into the target tissue and to set up a morphogen gradient; and also that HSPG-bound Fgf8 contributes to the minor slow-moving Fgf8 fraction and this interaction affects the signalling of Fgf8.

A morphogen functions by providing positional information through formation of a protein concentration gradient^{1,2}. Here we directly quantified the extracellular Fgf8 protein concentration gradient from a restricted source in gastrulating zebrafish embryos between sphere and germ-ring stage (Figs 1a and 2c, Supplementary Fig. 6a, b and Supplementary Information). The protein concentration decreased with increasing distance from the source, confirming that Fgf8–EGFP did establish a gradient in its target tissue (Fig. 4a). In contrast, secEGFP concentration remained nearly constant over distance (Fig. 4a).

As Fgf8–EGFP has a similar diffusion coefficient to that of secEGFP (Fig. 3b), this means that both molecules would be capable of spreading throughout the developing embryos by gastrulation. Concentration measurements taken at different developmental stages showed that secEGFP reached uniform distribution from sphere stage onwards, whereas Fgf8–EGFP maintained its gradient from the same stage (Supplementary Fig. 7 and Supplementary Information). Therefore, it was not the diffusion that limited the availability of the morphogen molecules throughout the tissue. Rather, only the presence of a sink for Fgf8 in the target tissue could sustain the shape of its gradient.

The observed Fgf8 gradient can be described by a simple model assuming a localized source and equilibrium between diffusion and degradation (sink) as expressed in the following reaction-diffusion equation¹⁰:

$$\frac{\partial C(\vec{r}, t)}{\partial t} = D \nabla^2 C(\vec{r}, t) - kC(\vec{r}, t) \quad (1)$$

in which $C(\vec{r}, t)$ is the concentration of the signalling molecule at position \vec{r} , t is the time, ∇^2 is the Laplace operator, D is the diffusion coefficient and k is the degradation rate. The localized source creates a current j across the source boundary.

We solved equation (1) for the particular geometry of the embryos (Supplementary Fig. 8a and Supplementary Information). The steady-state solution describes the Fgf8 concentration gradient very well (Supplementary Equation 12 and Supplementary Fig. 8b). The steepness of the decay of protein concentration against the distance from source can be described by the decay length λ (Supplementary Equation 12). By globally fitting the measured data from 40 embryos, we obtained $\lambda = (197 \pm 7) \mu\text{m}$, corresponding to an average of around nine cell diameters. At this distance, the concentration of Fgf8–EGFP had decayed by approximately 50% from its source (Fig. 4a, Supplementary Fig. 8b and Supplementary Information).

Taking the value of the diffusion coefficient D determined in FCS, we could calculate the half-life time $t_{0.5}$ of Fgf8–EGFP to be about

9 min ($t_{0.5} = \ln(2)/k$ and $k = D/\lambda^2$; Supplementary Information). In contrast, secEGFP had a much longer half-life of 84 min. Furthermore, we determined the half-life for Cy5-labelled recombinant Fgf8 protein by monitoring the temporal decrease of its concentration after injection into the extracellular space of zebrafish embryos (Methods, Supplementary Fig. 9a, b and Supplementary Information). The measured half-life of Fgf8–Cy5 (18 min) was comparable to that calculated for Fgf8–EGFP, confirming that we can estimate the Fgf8–EGFP half-life using the diffusion-degradation model (equation (1) and Supplementary Information).

One mechanism for generating a sink for Fgf8 is by receptor-mediated endocytosis in its target cells. Changing the level of endocytosis should therefore influence the shape of the concentration gradient. To test this, we first overexpressed a dominant-negative version of Dynamin (dnDYN) in the embryos. Dynamin is a GTPase required for the scission of endocytic vesicles from cell membranes²⁹. Endocytosis of Fgf8–EGFP was reduced in dnDYN-overexpressing embryos (Supplementary Fig. 10a–c), and the gradient of Fgf8–EGFP became much shallower compared to that of the control embryos ($\lambda_{\text{dnDYN}} = (271 \pm 29) \mu\text{m}$; Fig. 4b, c). When we overexpressed cyan fluorescent protein (CFP)-labelled RAB5c, a GTPase required for the fusion of endocytic vesicles with early endosomes²⁹, the amount of Fgf8–EGFP inside the target cells was increased (Supplementary Fig. 10a–c), and the gradient of Fgf8–EGFP became steeper ($\lambda_{\text{RAB5c}} = (164 \pm 14) \mu\text{m}$; Fig. 4b, c). The decrease in the decay length in RAB5c-overexpressing embryos was significant but small, probably owing to other rate-limiting steps in the uptake of Fgf8 upstream or parallel to RAB5c. The diffusion coefficient of Fgf8–EGFP in the presence of dnDYN or RAB5c did not differ from the control ($D_{\text{Fgf8-EGFP dnDYN}} = (55 \pm 14) \mu\text{m}^2 \text{s}^{-1}$, $n = 34$ and $D_{\text{Fgf8-EGFP RAB5c}} = (53 \pm 14) \mu\text{m}^2 \text{s}^{-1}$, $n = 23$, mean \pm s.d.; Fig. 3b). The calculated half-life of free, extracellular Fgf8–EGFP was increased to 15 min in dnDYN-overexpressing embryos and decreased to 6 min in RAB5c ones. These results show that endocytosis of Fgf8 is important for maintaining and shaping its protein gradient in the target tissue, and confirm a previously suggested restrictive clearance mechanism¹⁸.

To conclude, by using FCS in a live embryo system, we directly measured the molecular dynamics of Fgf8 within its morphogen gradient. Our results support a simple mechanism, involving a localized source, Brownian diffusion through the extracellular space and a sink in the target tissue generated by receptor-mediated endocytosis, to form and maintain a morphogen gradient. This study shows the potential of FCS to obtain crucial data such as diffusion coefficients and degradation rates *in vivo*. These physical parameters can act as the baseline for further experimental and theoretical studies of morphogen gradients.

METHODS SUMMARY

mRNA injections. pCS2+ plasmids containing *fgf8-egfp*, *secgfp*, *mrpf-gpi*, *zfd1-mrpf* and *rab5c-cfp* were linearized with NotI and transcribed with SP6 RNA polymerase (Ambion). pcDNA 3.1(–) plasmid containing *dnodyn* was linearized with AclI and transcribed with T7 RNA polymerase (Ambion). mRNA was

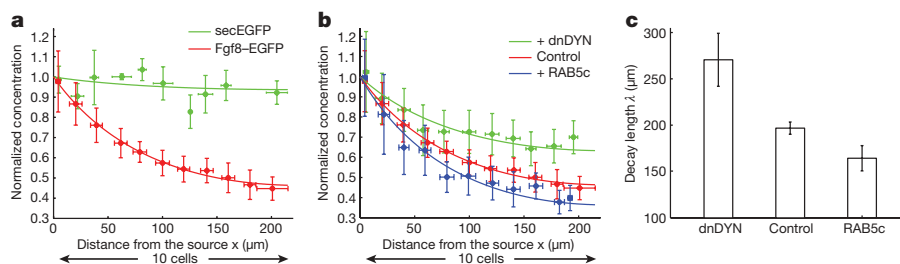


Figure 4 | Endocytosis regulates extracellular Fgf8 concentration gradient. **a**, Fgf8–EGFP concentration decreases as the distance from its source increases, whereas that of secEGFP remains constant in gastrulating zebrafish embryos. Data are binned in 20-μm intervals and fit with a radial model (Supplementary Information). Error bars denote s.d. **b**, The

Fgf8–EGFP concentration gradient becomes shallower in dnDYN-injected embryos ($n = 13$) and steeper in *rab5c-cfp*-injected embryos ($n = 12$), compared to the controls ($n = 40$). Error bars denote s.d. **c**, Comparison of the decay length λ in dnDYN, control and *rab5c-cfp*-injected embryos. Error bars denote the 95% confidence level.

injected into the cytoplasm of one-cell-stage embryos for overexpression of dnDYN (100 pg mRNA) and RAB5c-CFP (400 pg mRNA), and into the cytoplasm of a single cell of 32-cell-stage embryos for blastomere expression of Fgf8–EGFP (2 pg mRNA) and secEGFP (1 pg mRNA; Fig. 2b).

Whole-mount immunostaining. Antibody staining was carried out as described previously³⁰. Primary antibody: anti-myc (9e10)¹⁸ 1:25; secondary antibody: goat-anti-mouse Alexa-Fluor-488-conjugated (Invitrogen) 1:1,000.

Western blot. Embryos were collected between dome and 50% epiboly stages and manually de-yolked. The cells were dissociated, spun down and the supernatant containing the extracellular space proteins was collected for western blot³⁰. Primary antibody: anti-GFP (Santa Cruz) 1:100; secondary antibody: goat-anti-rabbit HRP-conjugated (Sigma) 1:3,000.

FCS. Embryos were mounted with animal pole oriented to the top and vegetal pole to the bottom in 1% low melting point agarose for FCS measurements. One-focus FCS was performed using Zeiss ConfoCor2 system with a C-Apochromat ×40, numerical aperture (NA) 1.2 water immersion objective. Two-focus FCS was performed on a home-built EMCCD-camera-based FCS setup (see Methods and Supplementary Information). One- or two-focus FCS measurements were taken in planes approximately 15 µm below the enveloping layer of the embryos. Measurements were performed at room temperature of 22 °C. Data were analysed using a home-written Matlab program (MathWorks), based on a weighted nonlinear-least-square-fitting routine. Irregular curves resulting from major instabilities were identified by distortions of the FCS curves and a systematic change in the intensity trace, and were excluded from further analysis.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 May; accepted 7 August 2009.

Published online 9 September 2009.

- Wolpert, L. Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* **25**, 1–47 (1969).
- Tabata, T. & Takei, Y. Morphogens, their identification and regulation. *Development* **131**, 703–712 (2004).
- Crick, F. Diffusion in embryogenesis. *Nature* **225**, 420–422 (1970).
- Kerszberg, M. & Wolpert, L. Mechanisms for positional signalling by morphogen transport: a theoretical study. *J. Theor. Biol.* **191**, 103–114 (1998).
- Entchev, E. V., Schwabedissen, A. & Gonzalez-Gaitan, M. Gradient formation of the TGF-β homolog Dpp. *Cell* **103**, 981–992 (2000).
- Strigini, M. & Cohen, S. M. Wingless gradient formation in the *Drosophila* wing. *Curr. Biol.* **10**, 293–300 (2000).
- McDowell, N., Gurdon, J. B. & Grainger, D. J. Formation of a functional morphogen gradient by a passive process in tissue from the early *Xenopus* embryo. *Int. J. Dev. Biol.* **45**, 199–207 (2001).
- Gregor, T., Bialek, W., de Ruyter van Steveninck, R. R., Tank, D. W. & Wieschaus, E. F. Diffusion and scaling during early embryonic pattern formation. *Proc. Natl Acad. Sci. USA* **102**, 18403–18407 (2005).
- Lander, A. D. Morphheus unbound: reimagining the morphogen gradient. *Cell* **128**, 245–256 (2007).
- Kicheva, A. et al. Kinetics of morphogen gradient formation. *Science* **315**, 521–525 (2007).
- Gregor, T., Wieschaus, E. F., McGregor, A. P., Bialek, W. & Tank, D. W. Stability and nuclear dynamics of the bicoid morphogen gradient. *Cell* **130**, 141–152 (2007).
- Boldajipour, B. et al. Control of chemokine-guided cell migration by ligand sequestration. *Cell* **132**, 463–473 (2008).
- Reifers, F. et al. *Fgf8* is mutated in zebrafish acerebellar (*ace*) mutants and is required for maintenance of midbrain-hindbrain boundary development and somitogenesis. *Development* **125**, 2381–2395 (1998).
- Reifers, F., Walsh, E. C., Leger, S., Stainier, D. Y. & Brand, M. Induction and differentiation of the zebrafish heart requires fibroblast growth factor 8 (*fgf8/acerebellar*). *Development* **127**, 225–235 (2000).
- Goldfarb, M. Functions of fibroblast growth factors in vertebrate development. *Cytokine Growth Factor Rev.* **7**, 311–325 (1996).
- Fürthauer, M., Reifers, F., Brand, M., Thisse, B. & Thisse, C. Sprouty4 acts *in vivo* as a feedback-induced antagonist of FGF signalling in zebrafish. *Development* **128**, 2175–2186 (2001).
- Raible, F. & Brand, M. Tight transcriptional control of the ETS domain factors Erm and Pea3 by Fgf signalling during early zebrafish nervous system development. *Mech. Dev.* **107**, 105–117 (2001).
- Scholpp, S. & Brand, M. Endocytosis controls spreading and effective signalling range of Fgf8 protein. *Curr. Biol.* **14**, 1834–1841 (2004).
- Rigler, R., Mets, Ü., Widengren, J. & Kask, P. Fluorescence correlation spectroscopy with high count rate and low-background: analysis of translational diffusion. *Eur. Biophys. J.* **22**, 169–175 (1993).
- Bacia, K., Kim, S. A. & Schwille, P. Fluorescence cross-correlation spectroscopy in living cells. *Nature Methods* **3**, 83–89 (2006).
- Amaya, E., Musci, T. J. & Kirschner, M. W. Expression of a dominant negative mutant of the FGF receptor disrupts mesoderm formation in *Xenopus* embryos. *Cell* **66**, 257–270 (1991).
- Ries, J., Yu, S. R., Burkhardt, M., Brand, M. & Schwille, P. Modular scanning FCS quantifies receptor-ligand interactions in living multicellular organisms. *Nature Methods*. doi:10.1038/nmeth.1355 (2 August 2009).
- Petrásek, Z. & Schwille, P. Precise measurement of diffusion coefficients using scanning fluorescence correlation spectroscopy. *Biophys. J.* **94**, 1437–1448 (2008).
- Brinkmeier, M., Dörre, K., Stephan, J. & Eigen, M. Two beam cross-correlation: a method to characterize transport phenomena in micrometer-sized structures. *Anal. Chem.* **71**, 609–616 (1999).
- Dertinger, T. et al. Two-focus fluorescence correlation spectroscopy: a new tool for accurate and absolute diffusion measurements. *ChemPhysChem* **8**, 433–443 (2007).
- Hou, S., Maccarana, M., Min, T. H., Strate, I. & Pera, E. M. The secreted serine protease xHtrA1 stimulates long-range FGF signalling in the early *Xenopus* embryo. *Dev. Cell* **13**, 226–241 (2007).
- Baeg, G. H., Selva, E. M., Goodman, R. M., Dasgupta, R. & Perrimon, N. The Wingless morphogen gradient is established by the cooperative action of Frizzled and Heparan Sulphate Proteoglycan receptors. *Dev. Biol.* **276**, 89–100 (2004).
- Desai, U. R., Wang, H. & Linhardt, R. J. Substrate specificity of the heparin lyases from *Flavobacterium heparinum*. *Arch. Biochem. Biophys.* **306**, 461–468 (1993).
- Robinson, M. S., Watts, C. & Zerial, M. Membrane dynamics in endocytosis. *Cell* **84**, 13–21 (1996).
- Westerfield, M. *The Zebrafish Book; a Guide for the Laboratory use of Zebrafish (Danio Rerio)* (Univ. of Oregon Press, 2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Brand and Schwille laboratories for discussions; A. Picker and K. Heinze for help with initial FCS measurements; B. Lutz for discussions; A. Oates, C. Boekel, A. Picker and P. Scott for comments on the manuscript; M. Fischer and K. Sippl for fish care; and A. Machate and B. Borgonov for technical assistance. This work was supported by an HFSP network grant (050503-50) to M.Br. and P.S., and by the EU Endotrack grant (050503-52) to M.Br.

Author Contributions S.R.Y. and M.Bu. performed the experiments, analysed the data and wrote the paper. M.N., J.R. and Z.P. provided advice on experimental design, data evaluation and interpretation. S.S. provided protocols for initial FCS measurements. P.S. and M.Br. initiated and supervised the collaboration, designed the project, analysed the data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.S. (petra.schwille@biotec.tu-dresden.de) or M.Br. (michael.brand@biotec.tu-dresden.de).

METHODS

Optical setup. One-focus FCS was performed using Zeiss ConfoCor2 system with a C-Apochromat $\times 40$, NA 1.2 water immersion objective. EGFP and Alexa Fluor 488 were excited by a 488-nm line of an Ar-Ion laser, and Cy5 was excited by a 633-nm HeNe laser using powers 5 and 7 μW , respectively. The emission was collected using a 505–550-nm band-pass and a 560-nm long-pass filter, respectively. Measurement time was 10×10 s. The ConfoCor2 system was calibrated using Alexa Fluor 488 dye ($D = 414 \mu\text{m}^2 \text{s}^{-1}$ at 25°C)³¹ in solution to determine the focal volume parameters $\omega_0 = (0.19 \pm 0.01) \mu\text{m}$ and $S = (5.5 \pm 0.6)$ (mean \pm s.d., $n = 15$).

Two-focus FCS was performed on a home-built FCS setup³², based on an inverted microscope (IX71, Olympus) with an UPlanApo $\times 60$, NA 1.2 water immersion objective, a 488-nm laser line (Sapphire 488-20, Coherent), a dichroic mirror (LP 505) and a band-pass filter (HQ 535/70M, AHF). The laser beam was split by a beam splitter cube, each beam was focused by a lens telescope through an acousto optical modulator (AOM, EQ Photonics) and the two beams were reunited again by a beam splitter cube. The reflected beam entered the objective back focal plane at a slight angle such that the two focal volumes had a lateral distance of $R = 1.1 \mu\text{m}$ in the sample. Laser power in each of the two excitation foci was 21 μW measured before the objective. Fluorescence was detected by an electron-multiplying CCD camera (EMCCD, iXon+DU897, Andor Technology). The EMCCD was operated in the fast kinetic acquisition mode³² using a home-written program based on the LabVIEW (National Instruments) driver supplied by Andor. The trigger output of the CCD was fed into a fast timer-counter-board (National Instruments) providing trigger high and low times for the AOMs that switched the two lasers alternately. This avoided spatial cross-talk of the two overlapping Gaussian excitation volumes. The effective measurement time was 60 s for each focal volume (4×15 s). The total measurement time was $9 \times$ longer owing to the camera readout time gaps in the non-continuous fast kinetic acquisition mode³². Each single measurement consisted of 1,000 repetitions, using 146 fast kinetic exposures of 200 μs , stored on the CCD chip. Therefore, the last correlation data point was at 15 ms. Image sequences of the camera were evaluated using Matlab (MathWorks).

Fgf8–EGFP diffusion coefficient measurement. Two picograms of *fgf8-egfp* mRNA was injected into a single cell at 32-cell stage as shown in Fig. 2b. Diffusion coefficient of Fgf8–EGFP was measured in the extracellular space in the target tissue (Fig. 2c, arrowheads).

Fgf8–Cy5 diffusion coefficient measurement. Recombinant Fgf8 protein was produced from *E. coli* bacteria (M15, Qiagen) and labelled with Cy5 dye (Amersham). Polystyrene microbeads (45.0 μm , Polysciences) were soaked in 50 ng ml^{-1} Fgf8–Cy5 solution overnight before they were implanted at the animal pole of the zebrafish embryos at sphere stage. FCS was used to measure the diffusion coefficient of the labelled protein in the extracellular space around the bead.

Fgf8–EGFP concentration measurement. Two picograms of *fgf8-egfp* mRNA was injected into a single cell at the 32-cell stage as shown in Fig. 2b. The progeny of the injected cell forms a restricted source of the fluorescent protein in gastrulating embryos. Concentrations of Fgf8–EGFP were measured at different distances away from the source (Fig. 2c, arrowheads).

As a measure for concentrations at particular positions in the embryos, the mean photon count rate I of the FCS measurement was used. It was a more robust measure for concentration than the deduced particle number N , when the focal volume was distorted in *in vivo* measurements due to mismatch between the refractive index of water and that of the embryonic tissue. A calibration experiment using Alexa Fluor 488 in solution showed that the count rate I changed linearly with the concentration of fluorophores, and the count rate was less influenced by a suboptimal focal volume than the deduced particle number N (Supplementary Fig. 6a, b and Supplementary Information).

Fgf8–Cy5 half-life measurement. To measure its half-life, $\sim 0.2 \text{ ng}$ of the labelled protein was injected directly into the extracellular space between sphere and dome stage. The injection was done at five different positions in the embryo to ensure fast and even distribution of the protein throughout the embryo. After about 15 min, FCS was used to measure the concentration of the labelled protein at different positions in the extracellular space over time (Supplementary Fig. 9a, b and Supplementary Information).

Cell transplantation. Donor embryos were injected with 100 pg zebrafish *fgf8* mRNA or 200 pg *fgf8-egfp* mRNA together with Rhodamine-dextran (10,000 Da, Invitrogen) at one-cell stage as cell lineage tracer. Zebrafish Fgf8- or Fgf8–EGFP-overexpressing cells were transplanted at 30% gastrula stage into the animal pole of a wild-type host embryo. The embryos were fixed at shield stage for *in situ* hybridization³⁰ and dextran detection (Vectastain ABC kit).

31. Petrov, E. P., Ohrt, T., Winkler, R. G. & Schwiile, P. Diffusion and segmental dynamics of double-stranded DNA. *Phys. Rev. Lett.* **97**, 258101 (2006).

32. Burkhardt, M. & Schwiile, P. Electron multiplying CCD based detection for spatially resolved fluorescence correlation spectroscopy. *Opt. Express* **14**, 5013–5020 (2006).

Reactive oxygen species prime *Drosophila* haematopoietic progenitors for differentiation

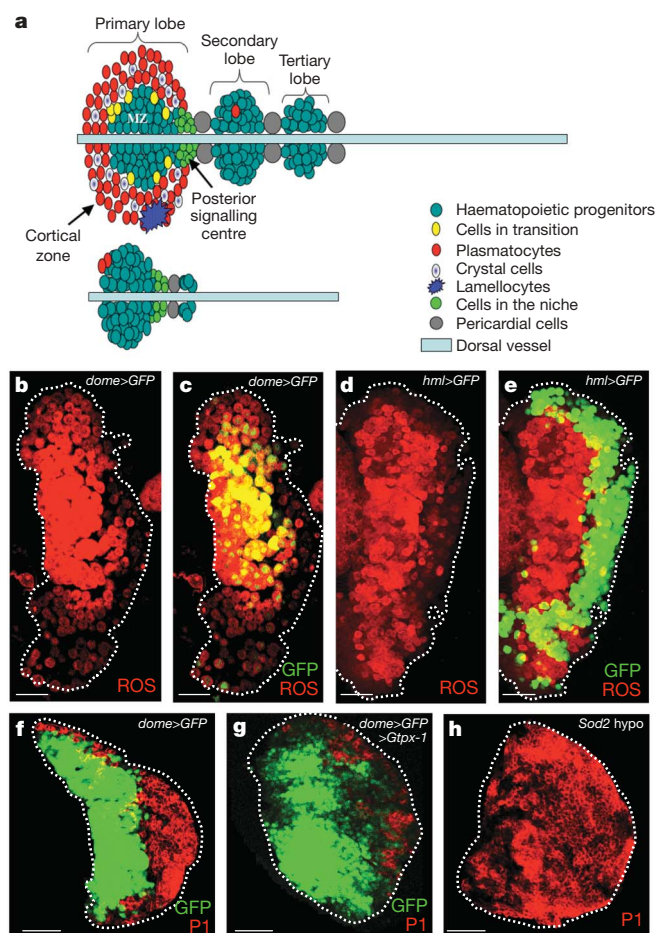
Edward Owusu-Ansah^{1†} & Utpal Banerjee^{1,2,3,4}

Reactive oxygen species (ROS), produced during various electron transfer reactions *in vivo*, are generally considered to be deleterious to cells¹. In the mammalian haematopoietic system, haematopoietic stem cells contain low levels of ROS. However, unexpectedly, the common myeloid progenitors (CMPs) produce significantly increased levels of ROS². The functional significance of this difference in ROS level in the two progenitor types remains unresolved^{2,3}. Here we show that *Drosophila* multipotent haematopoietic progenitors, which are largely akin to the mammalian myeloid progenitors⁴, display increased levels of ROS under *in vivo* physiological conditions, which are downregulated on differentiation. Scavenging the ROS from these haematopoietic progenitors by using *in vivo* genetic tools retards their differentiation into mature blood cells. Conversely, increasing the haematopoietic progenitor ROS beyond their basal level triggers precocious differentiation into all three mature blood cell types found in *Drosophila*, through a signalling pathway that involves JNK and FoxO activation as well as Polycomb downregulation. We conclude that the developmentally regulated, moderately high ROS level in the progenitor population sensitizes them to differentiation, and

establishes a signalling role for ROS in the regulation of haematopoietic cell fate. Our results lead to a model that could be extended to reveal a probable signalling role for ROS in the differentiation of CMPs in mammalian haematopoietic development and oxidative stress response.

The *Drosophila* lymph gland is a specialized haematopoietic organ which produces three blood cell types—plasmacytes, crystal cells and lamellocytes—with functions reminiscent of the vertebrate myeloid lineage^{5,6}. During the first and early second larval instars, the lymph gland comprises only the progenitor population (Fig. 1a, bottom diagram). However, by late third instar, multipotent stem-like progenitor cells become restricted to the medial region of the primary lymph gland lobe, in an area referred to as the medullary zone; whereas a peripheral

Figure 1 | ROS profile of third-instar lymph glands. **a**, Schematics of late third- (top) and early second-instar (bottom panel) lymph glands. Most progenitors are seen in the second instar and in the third instar, primary lobe central domain, the medullary zone. Differentiated cells are restricted to the cortical zone. Primary lobes are anterior, followed posteriorly by secondary and tertiary lobes. **b**, Medullary zone progenitors show increased ROS (red). Dotted outlines are based on high-power laser images. **c**, Expression of the medullary zone marker, *dome-gal4*, UAS-2×EYFP (green; genotype abbreviated '*dome>GFP*' for clarity) overlaps with the ROS dye (red) in medullary zone cells (therefore yellow). **d**, As **b**, the progenitors in the medullary zone show increased ROS levels (red). **e**, *hml¹-gal4*, UAS-2×EGFP is restricted to cells in the cortical zone (green). Most of the cells that are marked by *hml¹-gal4*, UAS-2×EGFP are low in ROS (therefore green) compared with cells in the medullary zone (red). A ring of *hml¹-gal4*, UAS-2×EGFP-expressing cells can be seen along the edge of the medullary zone that is both GFP and ROS positive (therefore yellow). These appear to be cells in a state of transition between a stem-like and a differentiated cell fate. **f**, P1 expression (red) in plasmacytes in the cortical zone. Expression of *dome-gal4*, UAS-2×EYFP (green) is restricted to the medullary zone as cortical zone cells (red) downregulate this marker. **g**, Overexpression of the antioxidant protein (GTPx-1) in progenitors (genotype: *dome-gal4*, UAS-2×EYFP; UAS-*Gtpx1*) causes pronounced reduction in the number of cells that express P1 (red). Some cells occupying the cortical zone region continue to express *dome-gal4*, UAS-2×EYFP whereas many others downregulate this marker without yet expressing the differentiation marker P1. **h**, In hypomorphic *Sod2/Sod2* homozygotes, in which a major ROS scavenger expression is reduced, P1 is expanded throughout the gland (red), not restricted to the cortical zone. This image represents the central domain. Scale bars, 50 μm.



¹Department of Molecular, Cell and Developmental Biology, ²Molecular Biology Institute, ³Department of Biological Chemistry, ⁴Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, California 90095, USA. †Present address: Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.

zone, referred to as the cortical zone, contains differentiated blood cells. By late third instar, the progenitors within the medullary zone are essentially quiescent, whereas the mature, differentiated population in the cortical zone proliferates extensively⁵. The posterior signalling centre is a group of about 30 cells (Fig. 1a, top diagram) that secretes several signalling molecules^{7–9} and serves as a stem-cell niche regulating

the balance between cells that maintain 'stemness' and those that differentiate^{8,9}.

Although several studies have identified factors that regulate the differentiation and maintenance of *Drosophila* blood cells and the stem-like progenitor population that generates them^{8–11}, intrinsic factors within the stem-like progenitors are less explored. Our interrogation of these intrinsic factors is the central theme of this investigation. We observed that by the third instar, the progenitor population in the normal wild-type lymph gland medullary zone contains significantly increased ROS levels compared with their neighbouring differentiated progeny that express mature blood cell markers in the cortical zone (Fig. 1b–e). ROS are not increased during the earlier larval instars but increase as the progenitor cells become quiescent and subside as they differentiate (Fig. 1b–e). This first suggested to us that the rise in ROS primes the relatively quiescent stem-like progenitor cells for differentiation. We reduced ROS by expressing antioxidant scavenger proteins GTPx-1 (ref. 12) (Fig. 1f, g) or catalase (Supplementary Fig. 1), specifically in the

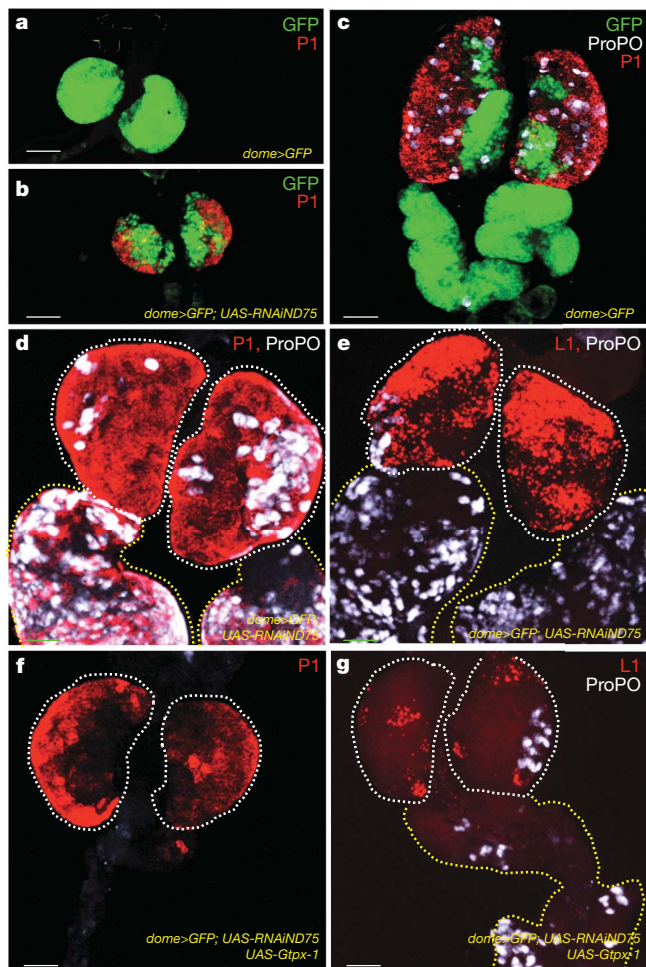


Figure 2 | Increased ROS production triggers precocious differentiation of the multipotent progenitors. In all panels, the progenitor population expresses the medullary zone marker, *dome-gal4, UAS-2×EYFP* (green). In **d–g**, the green channel has been omitted for clarity. The two genotypes used in these panels are control lymph glands (*dome-gal4, UAS-2×EYFP*), referred to as wild-type in this legend, and experimental lymph glands which express an RNAi construct to *ND75* (*dome-gal4, UAS-2×EYFP; UAS-RNAiND75*), abbreviated as *ND75 RNAi*. Scale bars, 50 μ m. **a**, P1 is not expressed (note absence of red) in early second-instar wild-type lymph glands. **b**, P1 expression (red) is robustly induced in early second-instar *ND75 RNAi* lymph glands. **c–e**, Disruption of *ND75* triggers precocious differentiation. **c**, In wild-type third-instar lymph glands, plasmotocytes marked with P1 (red) and crystal cells with ProPO (grey) are restricted to the cortical zone. These differentiated cell types are rarely, if ever, found in secondary and tertiary lobes. **d**, In third-instar *ND75 RNAi* lymph glands, there is a marked increase in P1- (red) and ProPO- (grey) expressing cells throughout the primary as well as the secondary and tertiary lobes (tertiary lobes are shown in Supplementary Fig. 4). **e**, Lamellocytes, marked by L1 (red) are prominently seen in third-instar *ND75 RNAi* lymph glands. Crystal cells are shown in grey. Lamellocytes are rarely found in secondary lobes. **f**, **g**, Scavenging ROS suppresses the differentiation associated with *ND75* disruption. Overexpression of Gtpx-1 in *ND75 RNAi* lymph glands (in **f** and **g**) potently suppresses differentiation into all three lineages as there is a decrease in P1 (red in **f**), ProPO (grey in **g**) and L1 (red in **g**) expression. Compare **f**, **g** with **d**, **e**. Controls for titration of GAL4 are shown in Supplementary Fig. 6.

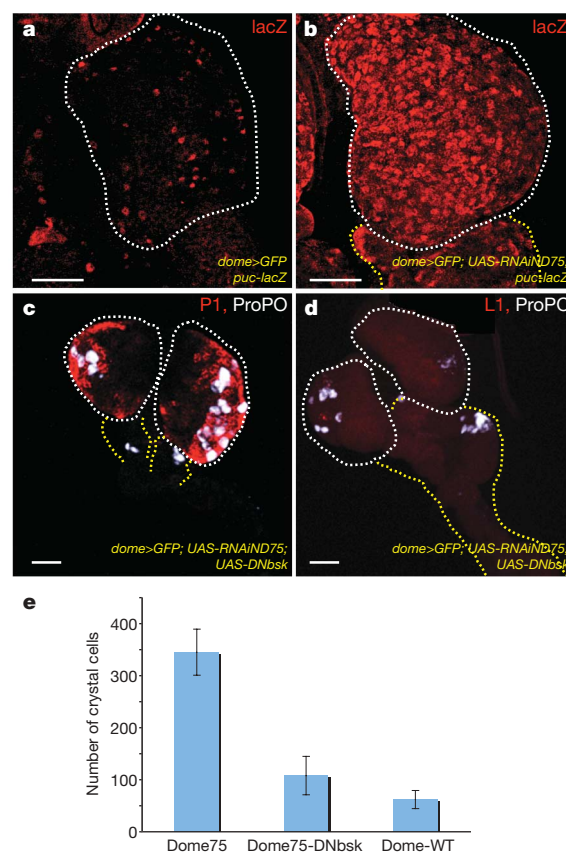


Figure 3 | Disrupting JNK signalling suppresses the ROS-dependent differentiation phenotype. The progenitor population also expresses the medullary zone marker, *dome-gal4, UAS-2×EYFP* (green), in **a–d**, but this has been omitted for clarity. Lymph glands that express an RNAi construct to *ND75* (*dome-gal4, UAS-2×EYFP; UAS-RNAiND75*), are abbreviated as *ND75 RNAi*. Scale bars, 50 μ m. **a**, **b**, JNK signalling is activated upon ROS increase. Expression of *puc-lacZ* (red) in wild-type lymph glands (**a**) and *ND75 RNAi* lymph glands (**b**). Levels of *puc-lacZ*, which is a transcriptional reporter of JNK signalling, are markedly increased in *ND75 RNAi* cells. **c**, **d**, JNK signalling is required for triggering differentiation associated with *ND75* disruption. Expressing a dominant negative construct of JNK in the precursor population ameliorates the effect of complex I disruption as the numbers of plasmotocytes (red in **c**), crystal cells (grey in **c** and **d**) and lamellocytes (red in **d**) are reduced virtually to wild-type levels. Compare **c**, **d** with Fig. 2d, e. **e**, Suppression of the number of crystal cells formed in *ND75 RNAi UAS-DNbsk* lymph glands relative to *ND75 RNAi* lymph glands. Error bars, s.e.m. ($n = 10$).

progenitor cell compartment using the GAL4/UAS system¹³, and found that suppressing increased ROS levels in haematopoietic progenitors significantly retards their differentiation into plasmacytes (Fig. 1f, g and Supplementary Fig. 1). As a corollary, mutating the gene encoding the antioxidant scavenger protein superoxide dismutase (*Sod2*)¹ led to a significant increase in differentiated cells and decrease in progenitors (Fig. 1h).

ROS levels in cells can be increased by the genetic disruption of complex I proteins of the mitochondrial electron transport chain^{14,15}, such as ND75 and ND42 (Supplementary Fig. 2). Unlike in wild type, where early second-instar lymph glands exclusively comprise undifferentiated cells (Fig. 2a), mitochondrial complex I depletion triggers premature differentiation of the progenitor population (Fig. 2b). This defect is even more evident in the third instar (compare Fig. 2c, d), where a complete depletion of the progenitors is seen as primary lobes are populated with differentiated plasmacytes and crystal cells. The third differentiated cell type, the lamellocyte, defined by the expression of the antigen L1, is rarely observed in the wild-type lymph gland (Supplementary Fig. 3) but is abundantly seen in the mutant (Fig. 2e). Finally, the secondary and tertiary lobes, largely undifferentiated in wild type, also embark on a robust program of differentiation upon complex I depletion (Fig. 2d, e and Supplementary Fig. 4). Importantly, the phenotype resulting from ND75 disruption can be suppressed by the co-expression of the ROS scavenger protein GTPx-1 (Fig. 2f, g; compare with Fig. 2d, e), which provides a causal link between increased ROS and the premature differentiation phenotype. Combining these results with those in Fig. 1, we conclude that the normally increased ROS levels in the stem-like progenitors serve as an intrinsic factor that sensitizes the progenitors to differentiation into all three mature cell types. Any further increase or decrease in the level of ROS away from the wild-type level enhances or suppresses differentiation respectively.

In unrelated systems, increased ROS levels have been demonstrated to activate the JNK signal transduction pathway^{1,16,17}. Consequently, we tested whether the mechanism by which the progenitors in the medullary zone differentiate when ROS levels increase could involve this pathway. The gene *puckered* (*puc*) is a downstream target of JNK signalling and its expression has been used extensively to monitor JNK activity¹⁸. Although *puc* transcripts are detectable by reverse transcriptase PCR (RT-PCR) (Supplementary Fig. 5), the *puc-lacZ* reporter is very weakly expressed in wild type (Fig. 3a). After disruption of ND75, however, a robust transcriptional upregulation of *puc-lacZ* expression can be seen (Fig. 3b), indicating that JNK signalling is induced in these cells in response to high ROS levels.

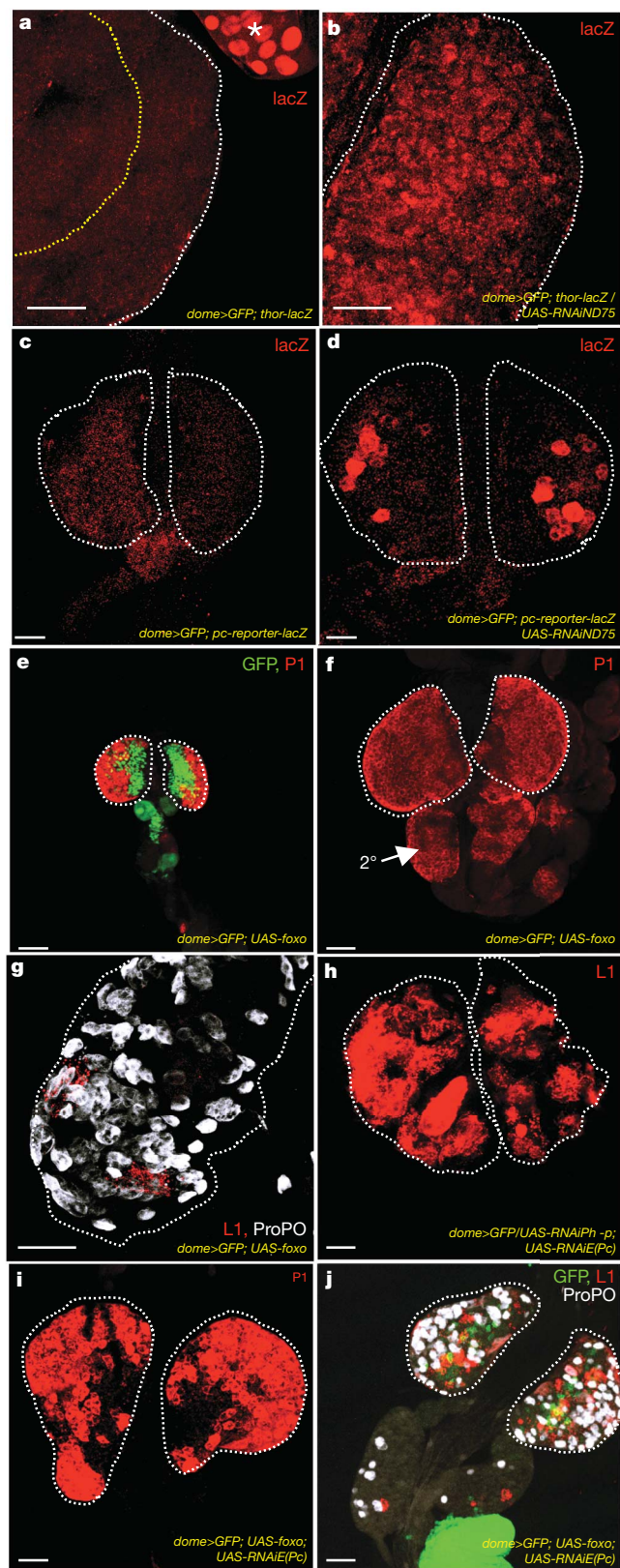


Figure 4 | FoxO activation and Polycomb downregulation phenocopy aspects of the ROS-induced differentiation. Progenitor cells express the medullary zone marker, *dome-gal4*, UAS-2×EYFP (green), omitted in some panels for clarity. Lymph glands from *dome-gal4*, UAS-2×EYFP larvae are referred to as wild-type controls in this legend. Cells expressing RNAi construct to ND75 in the progenitor cells (*dome-gal4*, UAS-2×EYFP; UAS-RNAiND75), are abbreviated as ND75 RNAi. Scale bars, 40 μm. **a, b**, Disruption of ND75 induces the FoxO reporter, *Thor-lacZ*. Wild type (**a**) does not express *Thor-lacZ* (absence of red). The asterisk points to *Thor-lacZ* expression in the ring gland, adjacent to the lymph gland. Expression of *Thor-lacZ* is significantly induced in ND75 RNAi lymph glands (**b**). **c, d**, Disruption of ND75 causes *polycomb* reporter expression. The *polycomb* reporter (red) is not expressed in wild type (**c**), but is induced in ND75 RNAi (**d**). **e–g**, FoxO overexpression causes increase in plasmacytes and crystal cells, but has no effect on lamellocytes. **e**, Overexpression of FoxO in the progenitors (*dome-gal4*, UAS-2×EYFP; UAS-foxo) causes premature differentiation into plasmacytes, evident as earlier than normal P1 staining (red) in a second-instar lymph gland (compare with Fig. 2a). **f**, Progenitors expressing FoxO in the medullary zone also initiate extensive differentiation into plasmacytes (red). Ectopic differentiation is also seen in secondary lobes (arrow, 2°). **g**, FoxO expression in the medullary zone causes increase in crystal cell number (grey). However, only a few isolated L1-positive cells (red) are evident. This image is at twice the magnification of other panels to highlight the few lamellocytes (red). **h**, RNAi-mediated downregulation of the expression of two polycomb proteins, Enhancer of polycomb, E(Pc), and polyhomeotic proximal, (Ph-p), leads to a robust increase in lamellocytes that stain for L1 (red). **i, j**, When FoxO and the RNAi construct to E(Pc) are expressed together in the medullary zone, there is an increase in all three mature cell markers. Co-expression of FoxO and an RNAi construct to E(Pc) trigger the full differentiation phenotype associated with complex I disruption as there is an increase in the number of plasmacytes (red in **i**), crystal cells (grey in **j**) and lamellocytes (red in **j**).

The precocious progenitor cell differentiation caused by mitochondrial disruption is suppressed upon expressing a dominant negative version of *basket* (*bsk*), the sole *Drosophila* homologue of JNK (Fig. 3c, d; compare with Fig. 2d, e; see also Supplementary Fig. 5). This suppression was associated with a decrease in the level of expression (Supplementary Fig. 5) of the stress response gene encoding phosphoenolpyruvate carboxykinase¹⁹; quantitatively a 68% suppression of the ND75 crystal cell phenotype was observed when JNK function was removed as well (Fig. 3e). Although disrupting JNK signalling suppressed differentiation, ROS levels remain increased in the mutant cells (Supplementary Fig. 2f), as would be expected from JNK functioning downstream of ROS.

In several systems and organisms, JNK function can be mediated by activation of FoxO as well as through repression of Polycomb activity^{17,20,21}. FoxO activation can be monitored by the expression of its downstream target *Thor*, using *Thor-lacZ* as a transcriptional read-out^{22–24}. This reporter is undetectable in wild-type lymph glands (Fig. 4a) although *Thor* transcripts are detectable by RT-PCR (Supplementary Fig. 5); however, the reporter is robustly induced when complex I is disrupted (Fig. 4b), which suggests that the increase in ROS that is mediated by loss of complex I activates FoxO. To monitor Polycomb de-repression, we used a *Polycomb* reporter, which expresses *lacZ* when Polycomb proteins are down-regulated. Although undetectable in wild-type lymph glands (Fig. 4c), disrupting ND75 leads to *lacZ* expression (Fig. 4d), which suggests that Polycomb activity is downregulated by the altered ROS and resulting JNK activation. Direct FoxO overexpression causes a remarkable advancement in differentiation to a time as early as the second instar (Fig. 4e), never seen in wild type (Fig. 2a). By early third instar, the entire primary and secondary lobes stained for plasmacyte (Fig. 4f) and crystal cell (Fig. 4g) markers when FoxO is expressed in the progenitor population. Unlike with ROS increase, we did not find a significant increase in lamellocytes upon FoxO overexpression. However, downregulating the expression of two polycomb proteins, Polyhomeotic Proximal (Php-x) and Enhancer of Polycomb (E(Pc)), that function downstream of JNK²¹ markedly increased lamellocyte number (Fig. 4h) without affecting plasmacytes and crystal cells (data not shown). When FoxO and a transgenic RNA interference (RNAi) construct against E(Pc) are expressed together in the progenitor cell population, differentiation to all three cell types is evident (Fig. 4i, j). We conclude that FoxO activation and Polycomb downregulation act combinatorially downstream of JNK to trigger the full differentiation phenotype: an increase in plasmacytes and crystal cells due to FoxO activation, and an increase in lamellocytes primarily due to Polycomb downregulation.

The analysis of ROS in the wild-type lymph gland highlights a previously unappreciated role for ROS as an intrinsic factor that regulates the differentiation of multipotent haematopoietic progenitors in *Drosophila*. Any further increase in ROS beyond the developmentally regulated levels, owing to oxidative stress, will cause the progenitors to differentiate into one of three myeloid cell types. It was reported previously² that the ROS levels in mammalian haematopoietic stem cells is low but that in the CMPs is relatively high. The *Drosophila* haematopoietic progenitors give rise entirely to a myeloid lineage and therefore are functionally more similar to CMPs than they are to haematopoietic stem cells. It is therefore a remarkable example of conservation to find that they too have high ROS levels. The genetic analysis makes it clear that the high ROS in *Drosophila* haematopoietic progenitors primes them towards differentiation. It will be interesting to determine whether such a mechanism operates in mammalian CMPs. In mice, as in flies, a function of FoxO is to activate antioxidant scavenger proteins. Consequently, deletion of FoxO increases ROS levels in the mouse haematopoietic stem cell and drives myeloid differentiation². However, even in the mouse haematopoietic system, FoxO function is dose and context dependent, as ROS levels in CMPs are independent of FoxO². Thus, although the basic logic of increased ROS in myeloid progenitors is conserved

between flies and mice, the exact function of FoxO in this context may have diverged.

Our past work¹⁴ and that of others^{1,25,26} has hinted that ROS can function as signalling molecules at physiologically moderate levels. This work supports and further extends this notion. Although excessive ROS is damaging to cells, developmentally regulated ROS production can be beneficial. The finding that ROS levels are moderately high in normal *Drosophila* haematopoietic progenitors and mammalian CMPs raises the possibility that wanton overdose of antioxidant products may in fact inhibit the formation of cells participating in the innate immune response.

METHODS SUMMARY

Lymph glands were stained as previously described^{5,8} using the following antibodies: mouse anti-P1 and L1, rat anti-ProPO, rabbit anti-β-gal (Cappel) and mouse anti-β-gal (Promega). Cy3-, Cy5- and FITC-conjugated secondary antibodies were from Jackson Laboratory. ROS staining was conducted as previously described^{14,15}. Images were captured using a BioRad Radiance 2000 confocal microscope with LaserSharp 2000 acquisition software.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 May; accepted 23 July 2009.

Published online 2 September 2009.

1. Finkel, T. & Holbrook, N. J. Oxidants, oxidative stress and the biology of ageing. *Nature* **408**, 239–247 (2000).
2. Tothova, Z. et al. FoxOs are critical mediators of hematopoietic stem cell resistance to physiologic oxidative stress. *Cell* **128**, 325–339 (2007).
3. Coffey, P. J. & Burgering, B. M. Stressed marrow: FoxOs stem tumour growth. *Nature Cell Biol.* **9**, 251–253 (2007).
4. Evans, C. J., Hartenstein, V. & Banerjee, U. Thicker than blood: conserved mechanisms in *Drosophila* and vertebrate hematopoiesis. *Dev. Cell* **5**, 673–690 (2003).
5. Jung, S. H., Evans, C. J., Uemura, C. & Banerjee, U. The *Drosophila* lymph gland as a developmental model of hematopoiesis. *Development* **132**, 2521–2533 (2005).
6. Wood, W. & Jacinto, A. *Drosophila melanogaster* embryonic haemocytes: masters of multitasking. *Nature Rev. Mol. Cell Biol.* **8**, 542–551 (2007).
7. Lebestky, T., Jung, S. H. & Banerjee, U. A. Serrate-expressing signaling center controls *Drosophila* hematopoiesis. *Genes Dev.* **17**, 348–353 (2003).
8. Mandal, L., Martinez-Agosto, J. A., Evans, C. J., Hartenstein, V. & Banerjee, U. A. Hedgehog- and Antennapedia-dependent niche maintains *Drosophila* haematopoietic precursors. *Nature* **446**, 320–324 (2007).
9. Krzemien, J. et al. Control of blood cell homeostasis in *Drosophila* larvae by the posterior signalling centre. *Nature* **446**, 325–328 (2007).
10. Lebestky, T., Chang, T., Hartenstein, V. & Banerjee, U. Specification of *Drosophila* hematopoietic lineage by conserved transcription factors. *Science* **288**, 146–149 (2000).
11. Bruckner, K. et al. The PDGF/VEGF receptor controls blood cell survival in *Drosophila*. *Dev. Cell* **7**, 73–84 (2004).
12. Missirlis, F. et al. A putative glutathione peroxidase of *Drosophila* encodes a thioredoxin peroxidase that provides resistance against oxidative stress but fails to complement a lack of catalase activity. *Biol. Chem.* **384**, 463–472 (2003).
13. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
14. Owusu-Ansah, E., Yavari, A., Mandal, S. & Banerjee, U. Distinct mitochondrial retrograde signals control the G1-S cell cycle checkpoint. *Nature Genet.* **40**, 356–361 (2008).
15. Owusu-Ansah, E., Yavari, A. & Banerjee, U. A protocol for *in vivo* detection of reactive oxygen species. *Nature Protocols*. doi:10.1038/nprot.2008.23 (2008).
16. Gotoh, Y. & Cooper, J. A. Reactive oxygen species- and dimerization-induced activation of apoptosis signal-regulating kinase 1 in tumor necrosis factor-α signal transduction. *J. Biol. Chem.* **273**, 17477–17482 (1998).
17. Essers, M. A. et al. FOXO transcription factor activation by oxidative stress mediated by the small GTPase Ral and JNK. *EMBO J.* **23**, 4802–4812 (2004).
18. Martin-Blanco, E. et al. puckered encodes a phosphatase that mediates a feedback loop regulating JNK activity during dorsal closure in *Drosophila*. *Genes Dev.* **12**, 557–570 (1998).
19. Harvey, K. F. et al. FOXO-regulated transcription restricts overgrowth of Tsc mutant organs. *J. Cell Biol.* **180**, 691–696 (2008).
20. Wang, M. C., Bohmann, D. & Jasper, H. JNK extends life span and limits growth by antagonizing cellular and organism-wide responses to insulin signaling. *Cell* **121**, 115–125 (2005).
21. Lee, N., Maurange, C., Ringrose, L. & Paro, R. Suppression of Polycomb group proteins by JNK signalling induces transdetermination in *Drosophila* imaginal discs. *Nature* **438**, 234–237 (2005).

22. Puig, O., Marr, M. T., Ruhf, M. L. & Tjian, R. Control of cell number by *Drosophila* FOXO: downstream and feedback regulation of the insulin receptor pathway. *Genes Dev.* **17**, 2006–2020 (2003).
23. Teleman, A. A., Hietakangas, V., Sayadian, A. C. & Cohen, S. M. Nutritional control of protein biosynthetic capacity by insulin via Myc in *Drosophila*. *Cell Metab.* **7**, 21–32 (2008).
24. Junger, M. A. *et al.* The *Drosophila* forkhead transcription factor FOXO mediates the reduction in cell number associated with reduced insulin signaling. *J. Biol.* **2**, 20 (2003).
25. Takeda, S. *et al.* Local positive feedback regulation determines cell shape in root hair cells. *Science* **319**, 1241–1244 (2008).
26. Foreman, J. *et al.* Reactive oxygen species produced by NADPH oxidase regulate plant cell growth. *Nature* **422**, 442–446 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank I. Ando and H. Muller for antibodies, and E. Hafen, A. Martinez-Arias, F. Missirlis, S. Noselli, R. Paro, S. Sinenko, the National Institute of Genetics Fly Stock Center (Japan) and the Bloomington Stock Center for fly stocks. We acknowledge M. Kulkarni and C. Pitsouli of the Perrimon laboratory for technical assistance. Owing to space limitations, we apologize to our colleagues whose work is not referenced. This study was supported by US National Institutes of Health grant R01HL067395 to U.B. and a T32 institutional postdoctoral fellowship T32-HL069766 to E.O.-A.

Author Contributions U.B. supervised the project. E.O.-A. conceived, designed and performed all experiments. E.O.-A. and U.B. discussed results and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to U.B. (banerjee@mbi.ucla.edu).

METHODS

Drosophila stocks and genetics. The following fly stocks were used: *domeless-gal4*, *hml^Δ-gal4* UAS-2xeGFP, UAS-GTPx1, UAS-cat, *puc^{ε69}*/TM3, Sb, the FLW-1 and LW-1 lines (for assessing polycomb downregulation). Oregon R, w; *P{tubP-GAL80^{ts}}*10; TM2/TM6B, Tb, *y¹ w^{67c23}*; *P{SUPor-P}Sod2^{KCG06854}*, UAS-2xeYFP on the X chromosomes, *y¹ w*; *P{lacW}Thor^{k13517}*, w; *P{UAS-bsk.K53R}20.1a* and *y¹ w**; *P{UAS-foxo.P}2* stocks are from the Bloomington stock centre. The RNAi stocks (UAS-ND75RNAi, UAS-ND42RNAi, UAS-E(Pc)RNAi and UAS-ph-pRNAi) were obtained from National Institute of Genetics Fly Stock Center (Japan). In all instances, crosses were set up at 25 °C for a day. For the third-instar RNAi experiments and all suppression experiments, the larvae were shifted to 18 °C for 2 days, and then transferred to 29 °C until dissection. For the second-instar experiments, larvae were transferred to 29 °C directly (without an intervening shift to 18 °C). For the overexpression experiments (involving UAS-foxo and

UAS-Gtpx-1) the larvae were maintained at 25 °C until dissection. In all instances, at least ten lymph glands were examined.

Gene expression analyses. Total RNA was extracted from third-instar lymph glands using TRIzol (Invitrogen), followed by RNA cleanup with the RNeasy Kit (Qiagen). The RNA QuantiTect Reverse Transcription Kit (Qiagen) was used for first-strand complementary DNA synthesis, and quantitative real-time PCR was performed with the QuantiTect SYBR Green PCR Kit (Qiagen) and analysed on an ABI Prism 7000 (Applied Biosystems) sequence detection system. The levels of actin transcripts were used to normalize total complementary DNA input. Relative quantification of transcript levels was calculated using the comparative *C_t* method. Results shown represent the mean of three replicates. For semi-quantitative RT-PCR, transcripts were analysed after the twentieth cycle, and actin transcripts served as loading controls. Primers used are shown in Supplementary Table 1.

LETTERS

A dimerization-dependent mechanism drives RAF catalytic activation

Thanashan Rajakulendran^{1,2*}, Malha Sahmi^{3*}, Martin Lefrançois³, Frank Sicheri^{1,2} & Marc Therrien^{3,4}

The ERK (extracellular signal-regulated kinase) pathway is an evolutionarily conserved signal transduction module that controls cellular growth, differentiation and survival¹. Activation of receptor tyrosine kinases (RTKs) by the binding of growth factors initiates GTP loading of RAS, which triggers the initial steps in the activation of the ERK pathway by modulating RAF family kinase function. Once activated, RAF participates in a sequential cascade of phosphorylation events that activate MEK, and in turn ERK. Unbridled signalling through the ERK pathway caused by activating mutations in RTKs, RAS or RAF has been linked to several human cancers². Of note, one member of the RAF family, BRAF, is the most frequently mutated oncogene in the kinase superfamily³. Not surprisingly, there has been a colossal effort to understand the underlying regulation of this family of kinases. In particular, the process by which the RAF kinase domain becomes activated towards its substrate MEK remains of topical interest. Here, using *Drosophila* Schneider S2 cells, we demonstrate that RAF catalytic function is regulated in response to a specific mode of dimerization of its kinase domain, which we term the side-to-side dimer. Moreover, we find that the RAF-related pseudo-kinase KSR (kinase suppressor of Ras) also participates in forming side-to-side heterodimers with RAF and can thereby trigger RAF activation. This mechanism provides an elegant explanation for the longstanding conundrum about RAF catalytic activation, and also provides an explanation for the capacity of KSR, despite lacking catalytic function, to directly mediate RAF activation. We also show that RAF side-to-side dimer formation is essential for aberrant signalling by oncogenic BRAF mutants, and identify an oncogenic mutation that acts specifically by promoting side-to-side dimerization. Together, our data identify the side-to-side dimer interface of RAF as a potential therapeutic target for intervention in BRAF-dependent tumorigenesis.

The function of KSR in RAF activation remains controversial. KSR contains a kinase domain of closest sequence similarity to RAF⁴ and was initially thought to drive RAF activation by virtue of its kinase activity. However, subsequent studies have been inconclusive in demonstrating intrinsic kinase activity for KSR and thus have relegated KSR to a class of pseudo-kinases⁵. Because of its capacity to bring MEK (also known as MAP2K) to RAF, the function of KSR is at present considered to be that of an organizing centre (or scaffold) in the ERK pathway⁶.

We previously showed in *Drosophila* S2 cells that co-overexpression of KSR with RAF and MEK stimulated RAF-dependent MEK phosphorylation⁷. If KSR was solely acting as a scaffold, we reasoned that overexpression of KSR without co-overexpression of its scaffold partners would perturb the optimal stoichiometry of KSR-containing complexes with the net effect of decreasing RAF activation. Because we observed increased RAF activation, this suggested that KSR might

possess an intrinsic RAF-activating capacity that becomes apparent on overexpression. If true, increasing levels of KSR should increasingly potentiate RAF activation. To test this possibility, we titrated in increasing amounts of KSR in S2 cells and monitored the effect by assessing RAF-dependent MEK phosphorylation⁸. As shown in Fig. 1, increasing levels of KSR correspondingly increased MEK phosphorylation. Surprisingly, we found that under these expression conditions KSR could drive robust RAF activation independently of other components normally required for RAF signalling in *Drosophila*. Specifically, we found that RNA interference (RNAi)-mediated knockdown of *ras*, *cnk* or *ave* (also known as *HYP*)^{7,9} had no effect on the ability of KSR to drive RAF activation (Fig. 1 and Supplementary Fig. 1a). Similarly, depletion of CK2, a kinase found to be recruited by KSR1 and which participates in BRAF and RAF1 activation¹⁰, had no effect on KSR-driven RAF activation (Supplementary Fig. 1b). Moreover, co-overexpression of a constitutively active RAS (RAS(Val 12)) barely augmented KSR-driven RAF activation under these conditions (Fig. 1). These results indicate that the overexpression of KSR in S2 cells unmasks an inherent activation potential on RAF beyond its well-established role as a scaffold, and bypasses the role of other components that impinge on RAF activation.

Recently, we identified a mutation in the kinase domain of KSR (Arg732His) that completely abolished its RAF activating capacity

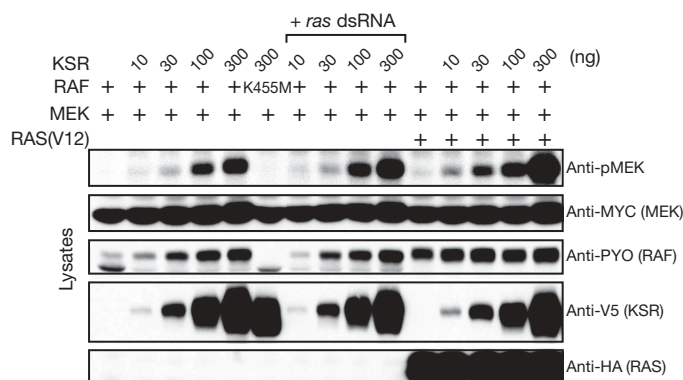


Figure 1 | KSR possesses intrinsic RAF activating potential. Co-overexpression of KSR, RAF and its substrate MEK as indicated in S2 cells leads to activation of RAF in a KSR concentration-dependent manner in the presence or absence of RNAi-mediated knockdown of *ras* or co-overexpression of a constitutively active RAS(Val12) variant (see Supplementary Fig. 9 for RNAi controls). RAF activation was assessed by immunoblotting for phosphorylated MEK (pMEK). The catalytically inactive RAF(Lys455Met) mutant served as a negative control. dsRNA, double-stranded RNA; HA, haemagglutinin; PYO, polyoma; V5, epitope from the RNA polymerase α subunit of simian virus 5.

¹Centre for Systems Biology, Samuel Lunenfeld Research Institute, Toronto, Ontario M5G 1X5, Canada. ²Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. ³Institute for Research in Immunology and Cancer, Laboratory of Intracellular Signaling, ⁴Département de pathologie et de biologie cellulaire, Université de Montréal, Montréal, Québec H3C 3J7, Canada.

*These authors contributed equally to this work.

and yet did not interfere with its ability to bring MEK to RAF⁷ (Supplementary Fig. 1c). Because the kinase domain of KSR is most similar to that of RAF⁴, we proposed that the previously determined crystal structure of the kinase domain of human BRAF might provide a good model to discern the mechanism of action of the KSR(Arg732His) mutation. Indeed, Arg 732 is not only invariant across all KSR proteins, but it is invariant across the larger KSR and RAF family (but not all kinases; Supplementary Fig. 2). Intriguingly, although the structure of the kinase domain of BRAF was reported as a monomer¹¹, the asymmetric unit of the crystal in fact contains two RAF kinase domains that interact in a unique side-to-side fashion involving the N-lobe of their kinase domains (Supplementary Fig. 3a). This mode of dimerization, which was not appreciated to date, can be observed in a total of five other RAF structure analyses^{11–14}, suggesting that the mode of dimerization is functionally relevant rather than an artefact of crystal packing (Supplementary Fig. 3b). Side-to-side dimerization of the RAF kinase domain provocatively engages helix α C, a key structural element whose conformation acts a regulatory function in several protein kinases¹⁵ (Supplementary Fig. 3a). Most notably, a specific mode of dimerization involving helix α C underlies an allosteric mechanism for kinase activation for both PKR¹⁶ and EGFR¹⁷ kinase domains (Supplementary Fig. 3c). As helix α C of the RAF kinase domain adopts a productive conformation in the dimeric crystal configuration¹¹, we reasoned that side-to-side dimerization itself might directly modulate the attainment of an active kinase conformation of RAF.

Projection of KSR and RAF conserved residues onto the RAF crystal structure showed that nearly the entire side-to-side dimer contact surface of RAF, but no other extended surfaces, are conserved across the larger KSR and RAF family (Fig. 2a). This suggested that KSR might form an analogous dimer structure. Moreover, the position of Arg 481 (the equivalent of Arg 732 in KSR; Supplementary Fig. 2) at the centre of the side-to-side dimer interface of the RAF crystal structure (Supplementary Fig. 4) indicated the basis by which the mutation of Arg 732 in KSR might exert a functional effect by perturbing dimerization (for simplicity, we use the *Drosophila* Raf (also known as Phl) numbering scheme for discussion of human BRAF positions; see Supplementary Table 1 for list of residue equivalence between BRAF and *Drosophila* Raf).

To investigate the potential of the RAF kinase domain to form dimers in solution, we performed analytical ultracentrifugation experiments. Equilibrium sedimentation analysis confirmed that RAF can form dimers, whereas mutation of Arg 481 at the dimer interface converted it to a predominant monomer in solution (Fig. 2b). Given the conserved nature of the dimer interface between RAF and KSR, we reasoned that the Arg732His mutation in KSR might similarly perturb the ability of KSR to form an analogous side-to-side homodimer or to form a side-to-side heterodimer with RAF. This in turn could explain how the KSR(Arg732His) mutation abolishes RAF activation.

If KSR mediates RAF activation by a mechanism involving the formation of a specific side-to-side homodimer with itself or a heterodimer with the kinase domain of RAF, then mutation of other dimer interface residues on KSR in addition to Arg 732 should also impair RAF activation. Using our minimal KSR–RAF–MEK co-overexpression activation assay, we found this to be the case. Specifically, individual mutation of four other residues (Gly700Trp, Phe739Ala, Met740Trp and Tyr790Phe) on KSR impeded its ability to induce RAF activation (Fig. 3a). Notably, none of the dimer interface mutations in KSR detectably affected the KSR–MEK interaction (not shown). If KSR mediates RAF activation by forming a specific side-to-side heterodimer with the kinase domain of RAF, then mutations of the corresponding residues on RAF (Gly450Trp, Phe488Ala, Met489Trp and Tyr538Phe) should equally impair RAF activation. As shown in Fig. 3a, we also found this to be the case. In contrast, control mutations remote from the dimer interface on the

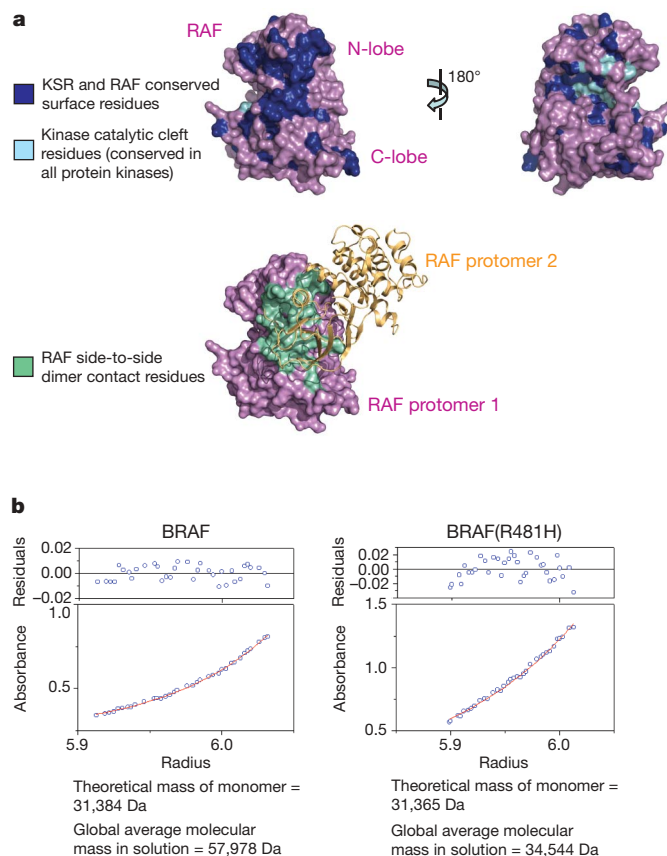


Figure 2 | KSR and RAF share a conserved side-to-side dimer interface.

a, Projection of highly conserved residues across both KSR and RAF orthologues onto the crystal structure (Protein Data Bank (PDB) accession 1UWH (ref. 11)) of the BRAF kinase domain (top panel) highlights common side-to-side dimer contact surfaces visualized originally in crystal structures of BRAF (bottom panel). **b**, Analytical ultracentrifugation analysis reveals that mutation of Arg 481 (Arg481His) in BRAF transitions the protein from a dimer (left panel) to a monomer (right panel) in solution. The red line denotes a fitted curve to the self-association model. The residuals for the fit are shown in the top panels.

kinase domains of both KSR and RAF showed no significant effect on RAF activation (Fig. 3a). These results confirm that the integrity of the side-to-side dimer interface on KSR and on RAF is essential for RAF activation.

To demonstrate that the formation of side-to-side kinase domain heterodimers by KSR and RAF *per se* leads to RAF activation, we used the FRB and FKBP fusion protein system to inducibly promote KSR–RAF side-to-side heterodimer formation by the addition of rapamycin *in vivo*¹⁸. Towards this end, we fused a region encompassing the minimal kinase domains of KSR and RAF to the FRB and FKBP fragments, respectively⁸ (see Supplementary Fig. 5 for schematic of assay). In this set-up, we observed that promoting a heterodimer between KSR and RAF by addition of rapamycin was indeed sufficient to potently activate RAF as evidenced by the increased levels of phosphorylated MEK (Fig. 3b). RAF activation was selectively perturbed by ten specific mutations at the side-to-side dimer interface on KSR and on RAF, but not by control mutations outside the interface (Fig. 3b and Supplementary Fig. 5). These results indicate that formation of the side-to-side heterodimer between KSR and RAF kinase domains is necessary for RAF activation under the conditions tested.

If RAF activation and downstream signalling is solely dependent on forming the kinase domain side-to-side dimer, then we wondered whether RAF kinase domain homodimers (in contrast to KSR–RAF heterodimers) might also promote RAF activation. To investigate this possibility, we used the FRB–FKBP–rapamycin system to drive

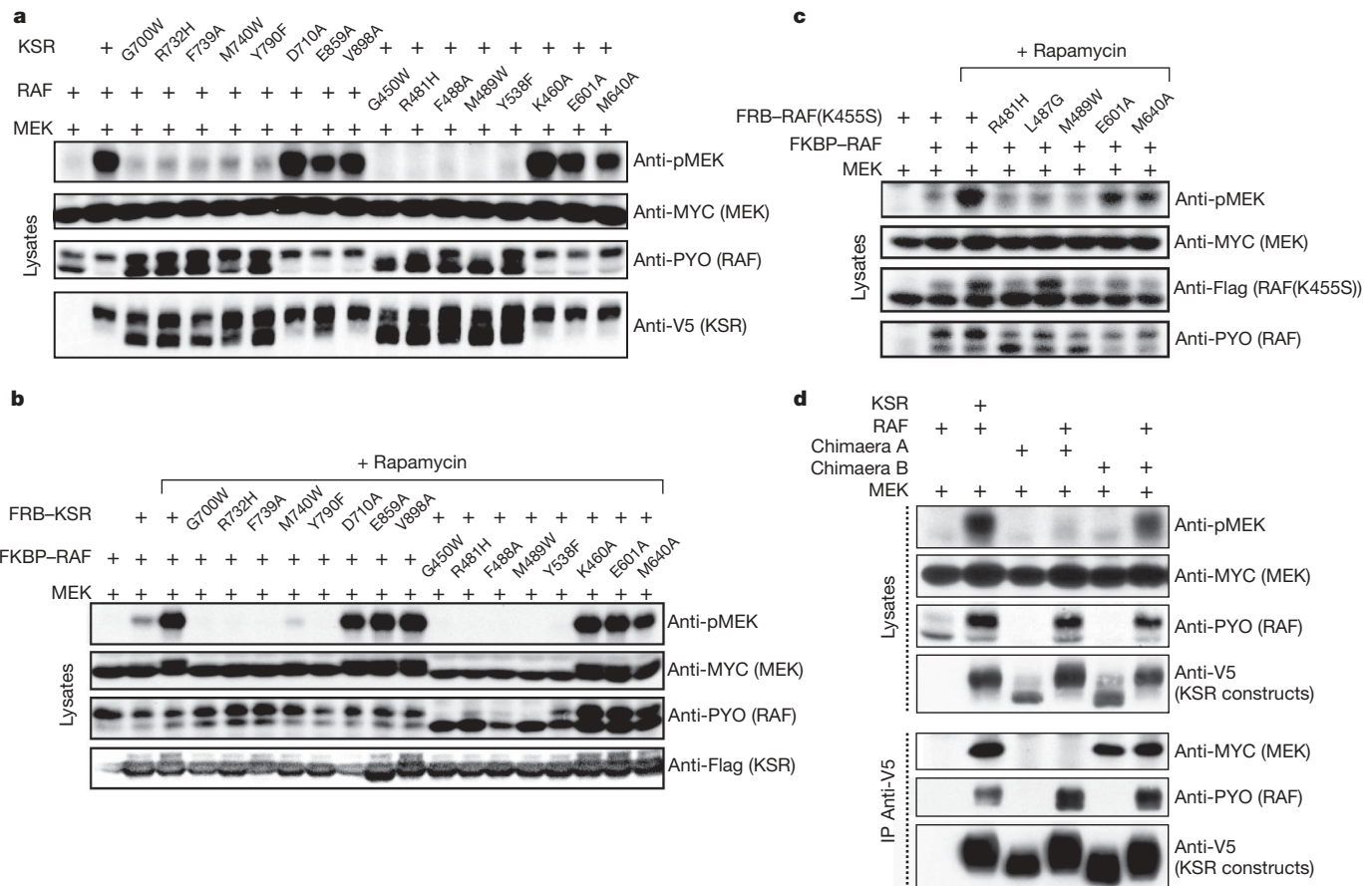


Figure 3 | Side-to-side dimerization of RAF activates its catalytic function. **a**, The individual effect of side-to-side dimer interface mutations on KSR and on RAF and their effect on RAF activation was assessed by monitoring the levels of phosphorylated MEK in S2 cells as performed in Fig. 1. Control mutations outside the interface correspond to Lys460Ala, Glu601Ala and Met640Ala in RAF, and Asp710Ala, Glu859Ala and Val898Ala in KSR. **b**, The RAF activation potential of the FRB or FKBP fused kinase domains of KSR

side-to-side homodimer formation of RAF kinase domains *in vivo*. To ensure that dimer formation-induced activation is not confounded by transautophosphorylation, we introduced a mutation (Lys455Ser) in the FRB-RAF fusion to impair its kinase activity. As shown in Fig. 3c, rapamycin induced formation of RAF homodimers can indeed drive RAF activation in a manner dependent on the ability to form the side-to-side dimers. Interestingly, although RAF homodimers are competent for activation, the extent of MEK phosphorylation is not as robust as that resulting from KSR-RAF heterodimers. We tested whether this difference might be due to the specific ability of KSR (but not RAF) to stably associate with MEK⁸. For this, we tested two full-length KSR-RAF chimaeras that had opposite MEK binding ability (see Supplementary Fig. 6 for schematic of chimaeric constructs). Chimaera A corresponded to the amino-terminal region of KSR fused to the catalytically compromised kinase domain of RAF(Lys455Ser), which does not stably associate with MEK, whereas chimaera B consisted of a KSR protein in which the kinase N-lobe had been replaced by the N-lobe of RAF(Lys455Ser) and maintained MEK binding (Fig. 3d). In contrast to chimaera A, chimaera B was nearly as potent as KSR in promoting MEK phosphorylation (Fig. 3d). These results indicate that KSR has at least two distinct functions whereby it first activates RAF by dimerization, and second KSR recruits MEK to facilitate downstream signalling. This second function seems to enhance KSR's RAF activation potential relative to RAF homodimers.

Together, our study indicates that dimerization of the RAF kinase domain with KSR or with other RAF molecules is central to its activation mechanism. We posit that other regulatory proteins that

and RAF (see Supplementary Fig. 5b for schematic of fusion constructs) were assessed by monitoring the levels of phosphorylated MEK in the presence or absence of rapamycin in S2 cells. **c**, Activation potential of FRB/FKBP-RAF homodimers was assessed as in **b**. **d**, The ability of wild-type KSR, RAF and KSR-RAF chimaeric constructs (as illustrated in Supplementary Fig. 6) to drive RAF activation in S2 cells was assessed by monitoring levels of phosphorylated MEK.

impinge on RAF activation, such as RAS, CNK and HYP, may also act by modulating dimerization. A case in point is the role of 14-3-3 proteins, which are known to activate RAF through promoting dimerization^{19–21}. Interestingly, the 14-3-3 consensus binding site in RAF is also conserved in KSR (Supplementary Fig. 7a), suggesting that 14-3-3 could also act to promote KSR-RAF heterodimers (Supplementary Fig. 7b). Consistent with this possibility, we found that depletion of endogenous 14-3-3 proteins perturbed KSR-dependent RAF activation, as did mutation of the consensus 14-3-3 binding site in KSR or RAF (Supplementary Fig. 7c).

The mapping of oncogenic mutations to the activation segment of BRAF proved unequivocally that the activation segment of RAF is a key modulator of its catalytic function^{11,22}. If dimerization is also a critical modulator of RAF catalytic function, we wondered whether certain oncogenic RAF mutations might promote kinase activity by promoting dimerization. We looked for oncogenic mutations that mapped to the side-to-side dimer interface and identified one such mutation, Glu558Lys (ref. 22), that promoted kinase domain dimerization in solution (Fig. 4a and Supplementary Fig. 8). To investigate how the Glu558Lys mutation functions to hyperactivate RAF *in vivo*, we used the FRB-FKBP-rapamycin system to assess RAF activation in S2 cells. When the Glu558Lys mutation was tested in a kinase-dead background, it strongly hyperactivated wild-type RAF *in trans* in a rapamycin-dependent manner (Fig. 4b). We reasoned that the added hydrogen-bonding potential of Glu558Lys with Ser 561 on the opposite protomer might promote dimerization (Fig. 4a). Consistent with this possibility, the hyperactivity of Glu558Lys *in trans* was selectively

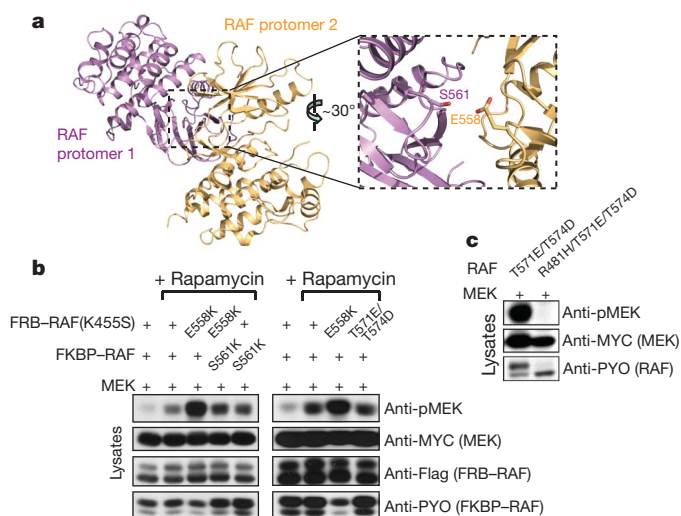


Figure 4 | Side-to-side dimer formation underlies the aberrant signalling potential of oncogenic RAF mutants. **a**, Glu 558 locates to the side-to-side dimer interface in RAF and is mutated to Lys in human cancers (Glu558Lys mutation; residue numbering scheme corresponds to *Drosophila* Raf). **b**, Catalytically inactive FRB-RAF(Lys455Ser) containing the Glu558Lys or Thr571Glu/Thr574Asp mutation was assessed for its activation potential towards FKBP-RAF and FKBP-RAF(Ser561Lys) *in trans* as in Fig. 3c. **c**, RAF activation assay using overexpressed full-length RAF and MEK proteins in S2 cells. The dimer interface mutation RAF(Arg481His) abrogates the pronounced activation potential of the activation segment mutation RAF(Thr571Glu/Thr574Asp).

attenuated towards the RAF(Ser561Lys) counterpart (Fig. 4b). These results indicate that the mechanism by which the oncogenic RAF(Glu558Lys) mutation acts is by promoting side-to-side dimers.

Given the *in trans* mechanism of action of the Glu558Lys mutation, we questioned whether the more prevalent class of oncogenic RAF mutations that impinge on modulation of the activation segment might also act to promote kinase domain dimerization. In contrast to the Glu558Lys mutant, the kinase-inactive version of a RAF activation segment gain-of-function mutant (RAF(Thr571Glu/Thr574Asp))⁷ did not show an enhanced ability to activate a wild-type counterpart *in trans* in our FRB-FKBP-rapamycin system (Fig. 4b). This suggests that the prevalent class of oncogenic mutations affecting the activation segment act exclusively *in cis* to promote intrinsic RAF activity. Consistent with the notion that dimerization is a critical modulator of RAF catalytic function, the hyperactivity of the RAF activation segment mutant *in cis* is ablated by a dimer interface mutation (Fig. 4c). These results raise the possibility that small molecule strategies directed at preventing the formation of side-to-side dimers by RAF could act as a therapeutic for RAF-dependent human tumours, one that would complement conventional strategies at present directed at inhibiting RAF enzymatic activity by blocking the catalytic cleft²³.

METHODS SUMMARY

The kinase domain fragment of recombinant human BRAF was expressed and purified from *Escherichia coli*. A structural homology model of *Drosophila* Ksr was generated using the structure of the kinase domain of human BRAF as a template. *Drosophila* S2 cells were used for all cell-based assays using *Drosophila* Raf. All mutants were generated using PCR and sequence verified. RAF activation was analysed in S2 cells by monitoring the RAF-mediated phosphorylation of MEK.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 April; accepted 24 July 2009.

Published online 2 September 2009.

- Wellbrock, C., Karasarides, M. & Marais, R. The RAF proteins take centre stage. *Nature Rev. Mol. Cell Biol.* **5**, 875–885 (2004).
- Roberts, P. J. & Der, C. J. Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene* **26**, 3291–3310 (2007).
- Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
- Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
- Boudeau, J., Miranda-Saavedra, D., Barton, G. J. & Alessi, D. R. Emerging roles of pseudokinases. *Trends Cell Biol.* **16**, 443–452 (2006).
- Kolch, W. Coordinating ERK/MAPK signalling through scaffolds and inhibitors. *Nature Rev. Mol. Cell Biol.* **6**, 827–837 (2005).
- Douziech, M., Sahmi, M., Laberge, G. & Therrien, M. A. KSR/CNK complex mediated by HYP, a novel SAM domain-containing protein, regulates RAS-dependent RAF activation in *Drosophila*. *Genes Dev.* **20**, 807–819 (2006).
- Roy, F., Laberge, G., Douziech, M., Ferland-McCollough, D. & Therrien, M. KSR is a scaffold required for activation of the ERK/MAPK module. *Genes Dev.* **16**, 427–438 (2002).
- Roignant, J. Y., Hamel, S., Janody, F. & Treisman, J. E. The novel SAM domain protein Aveugle is required for Raf activation in the *Drosophila* EGF receptor signaling pathway. *Genes Dev.* **20**, 795–806 (2006).
- Ritt, D. A. *et al.* CK2 is a component of the KSR1 scaffold complex that contributes to Raf kinase activation. *Curr. Biol.* **17**, 179–184 (2007).
- Wan, P. T. *et al.* Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell* **116**, 855–867 (2004).
- King, A. J. *et al.* Demonstration of a genetic therapeutic index for tumors expressing oncogenic BRAF by the kinase inhibitor SB-590885. *Cancer Res.* **66**, 11100–11105 (2006).
- Tsai, J. *et al.* Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antitumour activity. *Proc. Natl Acad. Sci. USA* **105**, 3041–3046 (2008).
- Hansen, J. D. *et al.* Potent and selective pyrazole-based inhibitors of B-Raf kinase. *Bioorg. Med. Chem. Lett.* **18**, 4692–4695 (2008).
- Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109**, 275–282 (2002).
- Dar, A. C., Dever, T. E. & Sicheri, F. Higher-order substrate recognition of eIF2 α by the RNA-dependent protein kinase PKR. *Cell* **122**, 887–900 (2005).
- Zhang, X., Gureasko, J., Shen, K., Cole, P. A. & Kuriyan, J. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell* **125**, 1137–1149 (2006).
- Muthuswamy, S. K., Gilman, M. & Brugge, J. S. Controlled dimerization of ErbB receptors provides evidence for differential signaling by homo- and heterodimers. *Mol. Cell Biol.* **19**, 6845–6857 (1999).
- Weber, C. K., Slupsky, J. R., Kalmes, H. A. & Rapp, U. R. Active Ras induces heterodimerization of cRaf and B-Raf. *Cancer Res.* **61**, 3595–3598 (2001).
- Garnett, M. J., Rana, S., Paterson, H., Barford, D. & Marais, R. Wild-type and mutant B-Raf activate C-Raf through distinct mechanisms involving heterodimerization. *Mol. Cell* **20**, 963–969 (2005).
- Rushworth, L. K., Hindley, A. D., O'Neill, E. & Kolch, W. Regulation and role of Raf-1/B-Raf heterodimerization. *Mol. Cell Biol.* **26**, 2262–2272 (2006).
- Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949–954 (2002).
- Halilovic, E. & Solit, D. B. Therapeutic strategies for inhibiting oncogenic BRAF signaling. *Curr. Opin. Pharmacol.* **8**, 419–426 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the F.S. and M.T. laboratories for advice and discussions. We are also grateful to R. Lambert and IRIC genomic platform for conducting qPCR reactions. This work was supported by grants from the Canadian Institutes for Health Research to F.S. (MOP-36399) and from the Canadian Cancer Society to M.T. (Award 018046). T.R. is a Research Fellow of The Terry Fox Foundation (Award 019684). F.S. is a recipient of a National Cancer Institute of Canada Scientist award, and M.T. is a Canada Research Chair (Tier II) in Intracellular Signaling.

Author Contributions T.R., M.S., F.S. and M.T. designed the study. T.R., M.S. and M.L. performed experiments. T.R. and M.S. analysed experiments. T.R., F.S. and M.T. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to F.S. (sicheri@lunenfeld.ca) or M.T. (marc.therrien@umontreal.ca).

METHODS

S2 expression plasmids. Copper-inducible pMet vectors were used for functional assays conducted in S2 cells as previously described^{7,8}. The FRB–RAF(Lys455Ser) fusion construct was assembled by inserting an *AseI/NotI* PCR fragment encompassing residues 328–738 of RAF into the *AseI/NotI* site of FRB–KSR⁸. The KSR–RAF chimaera A corresponds to KSR(1–665) fused to RAF(417–739), whereas chimaera B replaced the N-lobe of KSR (amino acid positions 666–757) with the one of RAF (amino acid positions 417–505). In both cases, the RAF N-lobe contained a Lys455Met change to catalytically impair its kinase activity and thereby mimicked kinase-inert KSR. Variant full-length *Drosophila* Ksr, Raf or FRB and FKBP fusion mutants were generated by QuickChange mutagenesis (Stratagene). Mutagenized complementary DNAs were fully sequenced to verify that only the desired mutations had been introduced.

S2 cell assays. S2 cells were maintained in serum-free insect cell medium (Sigma) at 27 °C. Cells were seeded at a density of 1.75×10^6 cells ml⁻¹ 24 h before transfection. Between 10 and 300 ng (or up to 900 ng for KSR(Arg732His)) of DNA was transfected per construct using Effectene (Qiagen). dsRNAs were produced and used in RNAi experiments as described⁸. Protein expression was induced by adding CuSO₄ (0.7 mM) 36 h before collecting the cells. For FRB and FKBP-mediated dimer formation, rapamycin (Sigma) was added (1 μM) to the medium 2 h before collecting the cells. Lysates, immunoprecipitations, western blot procedures and antibodies were essentially as previously described⁷.

Bacterial protein expression and purification. BRAF (residues 448–723) and mutant series (Arg481His, Leu487Arg and Leu487Arg/Glu558Lys) were recombinantly expressed from pETM-30 (EMBL) plasmid in *E. coli* BL21 cells as TEV protease-cleavable 6×His–glutathione *S*-transferase (GST)-tagged fusions. To increase the level of soluble protein expression in *E. coli*, 16 specific mutations (remote from the side-to-side dimer interface) were introduced in BRAF as described¹³. Expressed proteins were bound to Ni-NTA and eluted with imidazole and subjected to TEV protease treatment. Further purification was performed by subtractive Ni-NTA and size exclusion (Superdex 200) chromatography.

Homology modelling. A multiple sequence alignment of KSR and RAF kinase domains was used to build a structural model of the kinase domain of *Drosophila* Ksr (residues 670–945) in SWISS-MODEL²⁴. An initial model was generated using the structure of the kinase domain of BRAF as a template (chain A of PDB accession 1UWH (ref. 11)). This model was manually edited in COOT²⁵ and a poorly modelled loop spanning residues 821–838 was removed. To generate the KSR–RAF side-to-side heterodimer, the modelled structure of KSR was superimposed onto chain A of PDB entry 1UWH.

Analytical ultracentrifugation. Equilibrium sedimentation was performed with a Beckman Optima XL-A ultracentrifuge and An60Ti rotor. BRAF samples were

prepared in 20 mM Tris, pH 7.5, 200 mM NaCl, 5% glycerol and 1.5 mM TCEP for analysis. Data was collected at 4 °C for three protein concentrations (25, 12.5 and 6.25 μM) at three rotor speeds (13,000, 18,000 and 23,000 r.p.m. for wild-type BRAF and BRAF(Arg481His) or 12,000, 17,000 and 25,000 r.p.m. for BRAF(Leu487Arg) and BRAF(Leu487Arg/Glu558Lys)). Model analysis of the data was performed simultaneously in a global curve-fitting procedure (Origin software, Beckman). For this, data collected at 13,000 and 18,000 r.p.m. for wild-type BRAF was analysed at all three protein concentrations; data collected at 18,000 and 23,000 r.p.m. for BRAF(Arg481His) was analysed at all three protein concentrations; data collected at 17,000 and 25,000 r.p.m. for BRAF(Leu487Arg) was analysed at all three protein concentrations; data collected at 17,000 and 25,000 r.p.m. for BRAF(Leu487Arg/Glu558Lys) at 25 μM and 12.5 μM was analysed. The term ‘global’ refers to fits across all rotor speeds for a given concentration.

The global self association fit yielded an average molecular mass of 57,978 Da for wild-type BRAF. The ratio of the observed average molecular mass to the theoretical molecular mass of the monomer is 1.9:1, suggesting that the sample contains mostly dimers. A single-species dimer model (shown by the red line; Fig. 2b) best fit the observed data (blue circles; Fig. 2b), indicated by the random distribution of the residuals—a measure of goodness of fit (the residual is the difference between the observed value and the predicted value). For BRAF(Arg481His), the global self association fit yielded an average molecular mass of 34,544 Da. The ratio of the observed average molecular mass to the theoretical molecular mass of the monomer is 1.1:1, suggesting that the sample contains mostly monomers. A single-species monomer model (red line; Fig. 2b) best fit the observed data (blue circles; Fig. 2b). For BRAF(Leu487Arg), the global self association fit yielded an average molecular mass of 48,636 Da. The ratio of the observed average molecular mass to the theoretical molecular mass of the monomer is 1.5:1, suggesting that the sample contains a mixture of monomers and dimers. Consistent with this, a monomer-dimer model (red line; Supplementary Fig. 8) resulted in the best fit to the observed data (blue circles; Supplementary Fig. 8) with a dissociation constant (K_d) of ~2 μM. For BRAF(Leu487Arg/Glu558Lys), the global self association fit yielded an average molecular mass of 55,472 Da. The ratio of the observed average molecular mass to the theoretical molecular mass of the monomer is 1.8:1, suggesting that the sample contains mostly dimers and the data (blue circles; Supplementary Fig. 8) was best fit to a single-species dimer model (red line; Supplementary Fig. 8).

24. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* **31**, 3381–3385 (2003).

25. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

LETTERS

Active turnover modulates mature microRNA activity in *Caenorhabditis elegans*

Saibal Chatterjee¹ & Helge Großhans¹

MicroRNAs (miRNAs) constitute a large class of regulatory RNAs that repress target messenger RNAs to control various biological processes¹. Accordingly, miRNA biogenesis is highly regulated, controlled at both transcriptional and post-transcriptional levels², and overexpression and underexpression of miRNAs are linked to various human diseases, particularly cancers^{1,3}. As RNA concentrations are generally a function of biogenesis and turnover, active miRNA degradation might also modulate miRNA accumulation, and the plant 3'→5' exonuclease SDN1 has been implicated in miRNA turnover⁴. Here we report that degradation of mature miRNAs in the nematode *Caenorhabditis elegans*, mediated by the 5'→3' exoribonuclease XRN-2, affects functional miRNA homeostasis *in vivo*. We recapitulate XRN-2-dependent miRNA turnover in larval lysates, where processing of precursor-miRNA (pre-miRNA) by Dicer, unannealing of the miRNA duplex and loading of the mature miRNA into the Argonaute protein of the miRNA-induced silencing complex (miRISC) are coupled processes that precede degradation of the mature miRNA. Although Argonaute:miRNA complexes are highly resistant to salt, larval lysate promotes efficient release of the miRNA, exposing it to degradation by XRN-2. Release and degradation can both be blocked by the addition of miRNA target RNA. Our results therefore suggest the presence of an additional layer of regulation of animal miRNA activity that might be important for rapid changes of miRNA expression profiles during developmental transitions and for the maintenance of steady-state concentrations of miRNAs. This pathway might represent a potential target for therapeutic intervention on miRNA expression.

To identify and characterize an animal miRNA turnover pathway (Supplementary Fig. 1), we used the *let-7* miRNA as a model. *let-7* regulates cell fates in animals, and functions as a human tumour suppressor gene⁵. In *C. elegans*, the temperature-sensitive *let-7*(n2853) allele⁶ causes vulval bursting at the larval-to-adult transition when animals are grown at 25 °C. A point mutation in the mature miRNA impairs its binding to target mRNAs^{6,7}, and the level of the mutant miRNA is moderately decreased^{6,8} (Supplementary Fig. 2). Because decreased abundance seems to be functionally relevant⁷, we speculated that increased abundance of the mutant *let-7* and/or its 'sister' miRNAs *mir-48*, *mir-84* and *mir-241* (ref. 9) might suffice to downregulate some targets and partly suppress vulval bursting of *let-7*(n2853) animals. We examined diverse nucleases for their ability to suppress *let-7*(n2853)-associated lethality when depleted through RNA-mediated interference (RNAi) by feeding, initiated on synchronized L1 stage larvae. *xrn-2*, the orthologue of the yeast 5'→3' exoribonuclease Rat1p, potently suppressed vulval bursting, with more than 95% of animals surviving (Supplementary Figs 3 and 4, Supplementary Table 1 and Supplementary Text). By contrast, depletion of the *eri-1* ribonuclease, which degrades small interfering RNAs¹⁰, or *C. elegans* homologues of *Arabidopsis* SDN1 (ref. 4), did not suppress *let-7*(n2853) (Supplementary Table 1).

When we examined RNA from L4-stage *let-7*(n2853) worms, northern blotting revealed that *xrn-2*(RNAi) increased mature *let-7* levels relative to mock (empty vector) or *xrn-1*(RNAi) (Fig. 1a). This supports a role of XRN-2 in determining mutant *let-7* RNA accumulation.

Yeast Xrn2p/Rat1p performs transfer-RNA quality control, selectively removing incompletely modified tRNAs¹¹. However, the effects of depletion of *C. elegans* *xrn-2* extended to wild-type *let-7* and unrelated miRNAs (Fig. 1a and Supplementary Fig. 5), supporting a broad function of XRN-2 in miRNA homeostasis. Overexpression of *let-7* from a transgene¹² was modestly enhanced by *xrn-2*(RNAi) at an early stage (L3), but not at later developmental stages when transgene expression was already very high (Supplementary Fig. 6), suggesting that additional homeostatic mechanisms might curtail miRNA overexpression.

The *Arabidopsis* Xrn2p/Rat1p homologues XRN2 and XRN3 degrade the loop sequence of miRNA precursors without affecting levels of mature miRNAs¹³. To investigate whether *C. elegans* XRN-2 processed pre-miRNAs, we examined the abundant pre-*mir-60* (ref. 14). Pre-*mir-60* levels were unaltered by *xrn-2* depletion, whereas mature *mir-60* accumulated (Fig. 1b). Moreover, although depletion of the pre-miRNA processing enzyme Dicer (DCR-1)^{15,16} caused the accumulation of various low-abundance pre-miRNAs, *xrn-2*(RNAi) did not (Supplementary Fig. 5). Real-time polymerase chain reaction after reverse transcription (RT-qPCR) further revealed that the levels of the primary *let-7* (pri-*let-7*) and pri-*mir-77* transcripts remained unchanged on *xrn-2*(RNAi) (Fig. 1c). Thus, *xrn-2* preferentially, possibly exclusively, affects the accumulation of mature miRNAs. To our knowledge, this is the first example of XRN-2 functioning in the turnover of mature, fully functional RNA species.

To eliminate the possibility that XRN-2 simply cleared away inactive miRNAs instead of terminating miRNA activity, we examined the levels of *daf-12* and *lin-41* mRNAs, two *let-7* targets (refs 17, 18) that accumulate when *let-7* activity is lost^{8,19}. Depletion of *xrn-2* in *let-7*(n2853) worms decreased the abundance of *daf-12* and *lin-41* to wild-type levels (Fig. 1d), revealing a molecular basis for suppression of *let-7*(n2853) by *xrn-2*(RNAi). Thus, XRN-2 modulates *let-7* activity, rather than acting as a 'scavenger' enzyme.

To examine miRNA turnover biochemically, we developed an *in vitro* system using larval lysates and radiolabelled miRNAs. Initially, 3'-pCp-labelling or biotinylation were used to block the 3' end of *let-7* against 3'→5' exonucleolysis. Wild-type worm lysate converted these substrates to mononucleotides without the production of any visible intermediates at both 25 °C, the physiological temperature, and 37 °C (Fig. 1e and Supplementary Fig. 7). Because the latter temperature yielded more product, and for technical convenience, we performed subsequent reactions at 37 °C.

A 5'-labelled synthetic *let-7* and an internally labelled *let-7* *in vitro* transcript having free 3'-hydroxyl groups were similarly degraded when exposed to the lysate (Supplementary Fig. 8a, b). Thin-layer chromatography identified the product as nucleotide 5'-monophosphate

¹Friedrich Miescher Institute for Biomedical Research, PO Box 2543, CH-4002 Basel, Switzerland.

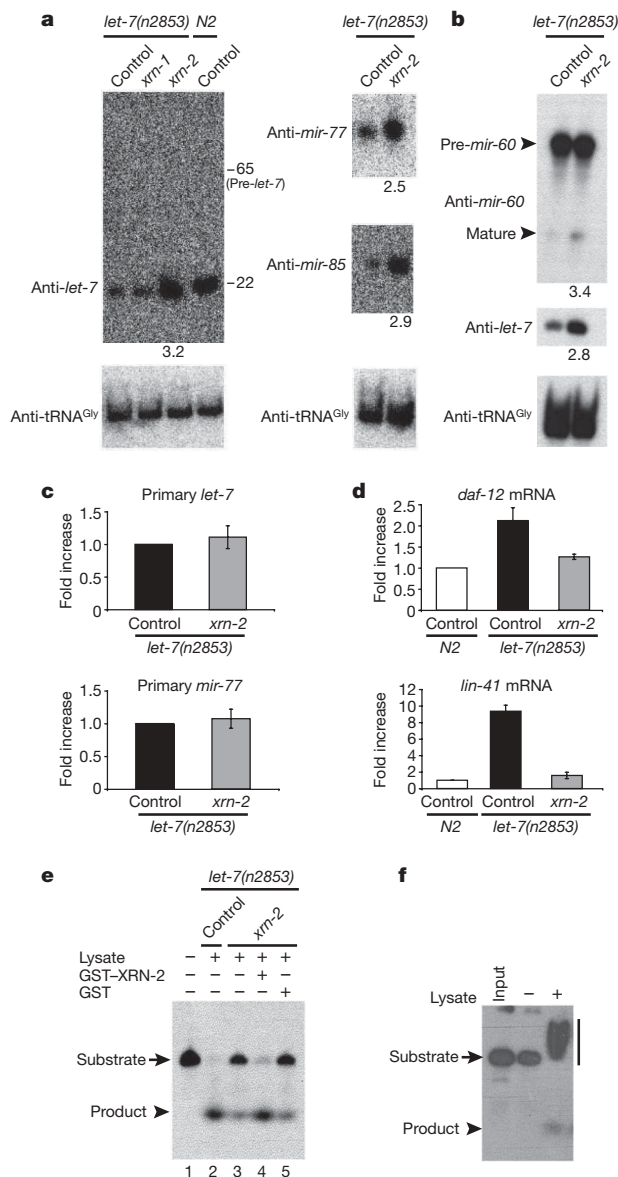


Figure 1 | Depletion of *xrn-2* increases mature miRNA levels and activity. **a, b**, Northern blotting of RNA from *let-7(n2853)* and wild-type (N2) worms on RNAi as indicated. The fold increase in mature miRNA normalized to tRNA levels is indicated below the relevant blots. **a**, *let-7*, *mir-77* and *mir-85*. Numbers to the right of the left blot indicate sizes in nucleotides. **b**, *mir-60*, its abundant precursor (pre-*mir-60*) and *let-7*. **c**, Pri-miRNA levels determined by RT-qPCR ($n = 3$; means \pm s.e.m.). **d**, Levels of *let-7* targets *lin-41* and *daf-12* determined by RT-qPCR ($n = 3$; means \pm s.e.m.). **e**, Incubation of 3'-pCP-labelled *let-7* with lysates as indicated, containing or lacking XRN-2. **f**, Incubation of 5'-end-labelled yeast tRNA^{Phe} with lysate yields an array of slow-migrating bands (vertical bar) and a final product of a few nucleotides in length (arrowhead).

(Supplementary Fig. 8c), revealing the expected hydrolytic mode of degradation²⁰, although 3'→5' hydrolytic exonucleases present in the lysate might have contributed to this pattern. Nuclease activity was largely sequence independent: four other synthetic miRNAs were similarly degraded (not shown). By contrast, tRNA, another small RNA with a free 3'-hydroxyl group, yielded a distinct turnover pattern (Fig. 1f). Thus, RNA structure rather than sequence might determine which RNA turnover pathway engages a given substrate in the lysate.

The exonuclease activity depended on XRN-2, because miRNA degradation was impaired with *xrn-2(RNAi)* but not *xrn-1(RNAi)* lysate (Fig. 1e and Supplementary Figs 7b and 9). Depletion of *xrn-2* was specific because the addition of bacterially expressed, recombinant glutathione S-transferase (GST)-tagged XRN-2 (Supplementary Fig. 10),

but not GST alone, restored miRNA turnover (Fig. 1e; compare lanes 4 and 5). Consistent with the preference of 5'→3' exonucleases for a 5' phosphate on their substrates²¹ was our observation that miRNA turnover was also slowed down when 5'-non-phosphorylated substrate was used (Supplementary Fig. 11). We can therefore attribute the miRNA decay activity acting in these lysates to XRN-2.

The *in vivo* assays indicated that XRN-2 affected the accumulation of the mature miRNA but not that of its precursors (Figs 1a and 2a, and Supplementary Fig. 5). To confirm this *in vitro*, we incubated radiolabelled, 5'-monophosphorylated pre-*let-7* with lysate. This substrate was converted into several products including mononucleotides (Fig. 2b), the sole product of mature miRNA turnover. When we repeated the assay with *dcr-1(RNAi)* lysate, we observed pre-*let-7* stabilization (Fig. 2c), as we did for the endogenous RNA *in vivo* (Supplementary Fig. 12). Product formation in the control lysate therefore depended on cleavage by Dicer.

In *xrn-2(RNAi)* lysate, pre-*let-7* still disappeared, which was consistent with the lack of pre-*let-7* accumulation on *xrn-2* depletion *in vivo*. However, a band migrating with a synthetic mature *let-7* accumulated in the *xrn-2(RNAi)* lysate but not in the mock RNAi lysate (Fig. 2d). We identified this band as mature *let-7* by northern analysis when performing the assay with unlabelled substrate (Fig. 2e). Pretreatment of lysates with micrococcal nuclease removed all endogenous RNA, ensuring that the mature miRNA detected by northern analysis was derived exclusively from the exogenous pre-miRNA. Micrococcal nuclease activity was terminated by the addition of

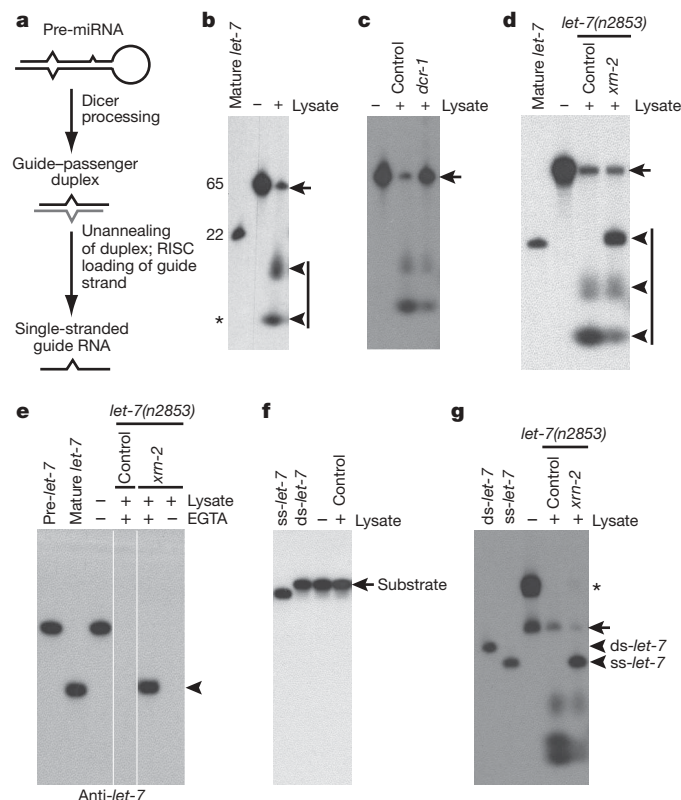


Figure 2 | Coordination of *in vitro* miRNA processing and turnover.

Internally radiolabelled, 5'-monophosphorylated pre-*let-7* transcript was used unless indicated otherwise. **a**, Outline of pre-miRNA processing. **b**, Pre-*let-7* incubated with lysate. Sizes in nucleotides are indicated at the left; asterisk: gel running front. **c**, Pre-*let-7* is stabilized in *dcr-1(RNAi)* lysates. **d**, *In situ*-diced *let-7* accumulates in *xrn-2(RNAi)* lysate. **e**, Turnover of unlabelled pre-*let-7* followed by northern blotting. All lanes are from a single autoradiograph. **f**, Native gel analysis of duplex *let-7* (5'-end-labelled guide) incubated with lysate. **g**, Native gel analysis of pre *let-7* turnover products. Right asterisk, pre-*let-7* conformer; arrows, substrates; arrowheads, reaction products; ss, single-stranded; ds, double-stranded.

EGTA before the substrate was added, and no signal was detected when EGTA was omitted, confirming efficient micrococcal nuclease activity (Fig. 2e). Pre-*let-7* cleavage by Dicer is therefore a prerequisite for miRNA degradation by XRN-2. Inefficient accumulation of mature miRNA was also observed with pre-*let-7* processing in *Drosophila* embryo lysate, leading to the speculation that mature miRNA might be unstable²². Thus, miRNA turnover activity might be conserved in other animals.

The accumulation of an additional product about 10 nucleotides (nt) long (Fig. 2b–d and Supplementary Fig. 13) remained largely unaffected by *xrn-2(RNAi)* (Fig. 2d) but decreased on depletion of *dcr-1* (Fig. 2c) and might consist of partly degraded 'loop' RNA, generated by processing of the pre-miRNA through Dicer.

One interpretation of our results was that XRN-2 degraded the mature single-stranded miRNA in the guide–passenger RNA duplex that Dicer generates from the pre-miRNA (Fig. 2a). However, a synthetic guide–passenger duplex remained stable in lysate (Fig. 2f), suggesting that degradation should occur after separation of the two strands. Indeed, the mature *let-7* generated by pre-*let-7* processing in *xrn-2(RNAi)* lysates was single stranded, as revealed by native gel electrophoresis (Fig. 2g). The *in vitro* system therefore recapitulates several steps in miRNA biogenesis and turnover; that is, faithful pre-miRNA processing by Dicer, unannealing of the guide–passenger duplex, and degradation of the single-stranded miRNA.

Because single-stranded miRNA was degraded whereas a guide–passenger duplex was not, we speculated that target binding might modulate miRNA stability. To test this possibility we supplemented lysates with *in vitro*-transcribed *let-7* target RNA²³; that is, luciferase coding sequence fused to an artificial 3' untranslated region (UTR) containing three *let-7*-binding sites or control transcripts with mutated *let-7*-binding sites or lacking the 3' UTR entirely (Fig. 3a and Supplementary Fig. 14). Under these conditions, the transcript with *let-7*-binding sites, but not the control transcripts, stabilized the mature *let-7* miRNA produced from pre-*let-7* (Fig. 3b and Supplementary Fig. 15). The controls exclude the possibility that excess exogenous RNA simply quenched ribonuclease activity. A northern blot of an assay using unlabelled substrate confirmed mature *let-7* accumulation in the presence of target RNA (Fig. 3c). The effect was not restricted to *let-7*: identical results were obtained for pre-*mir-237* (Fig. 3d and Supplementary Figs 16 and 17). We conclude that targets can modulate the extent of mature miRNA degradation *in vitro*.

Because most cellular miRNA is thought to be associated with Argonaute proteins in the miRISC complex, we wished to confirm Argonaute-binding of the *in situ*-processed mature miRNAs. We therefore incubated radiolabelled pre-*let-7* with lysates from worms expressing green fluorescent protein (GFP)-tagged versions of both *C. elegans* miRNA argonautes, ALG-1 and ALG-2 (GFP/AGO)²⁴. When GFP/AGO was subsequently immunoprecipitated from lysate lacking miRNA target, no radiolabelled RNA, precursor or mature, was co-immunoprecipitated. By contrast, addition of *let-7* target RNA to the lysate permitted the co-immunoprecipitation of mature miRNA with GFP/AGO (Fig. 3e and Supplementary Fig. 18), demonstrating incorporation into miRISC.

Depletion of *xrn-2* caused substantial accumulation of mature *let-7* from pre-*let-7* in GFP/AGO larval lysates (Fig. 3e) but permitted the co-immunoprecipitation of only modest amounts of mature *let-7* with GFP/AGO, whereas abundant *let-7* remained in the post-immunoprecipitation supernatant (Fig. 3e and Supplementary Fig. 18). We conclude that, *in vitro*, miRNAs are dislodged from ALG-1/2 through a mechanism that is efficiently modulated by the target RNA binding status of the miRNA but is only partly dependent on XRN-2. Because both the miRNA 5' and 3' ends are thought to be bound directly by Argonaute²⁵, they would be inaccessible to exonucleases, necessitating additional factors for release. At the same time, release and degradation steps seem to be tightly coupled *in vivo*,

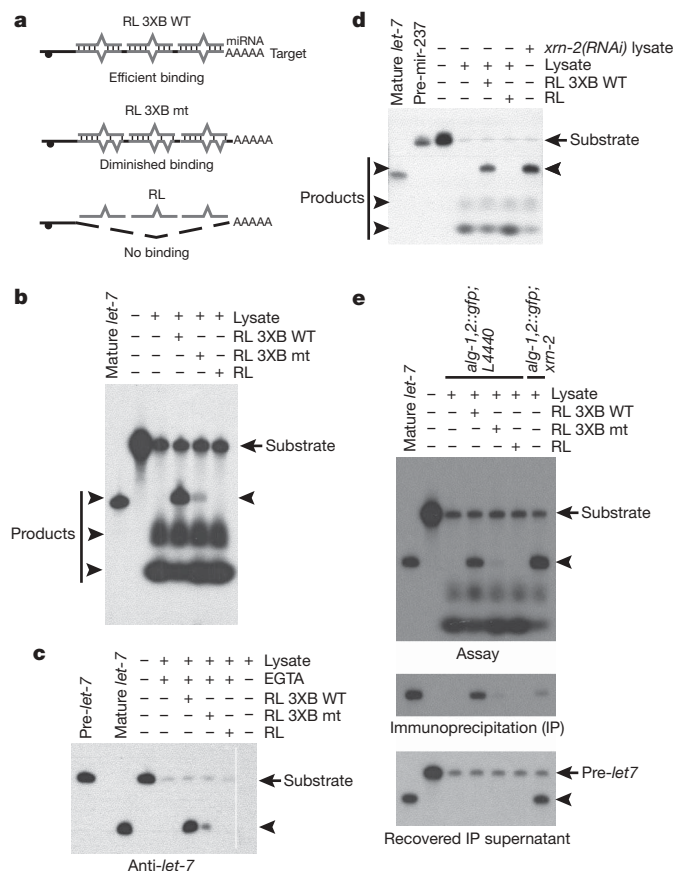


Figure 3 | Target-mediated stabilization of mature miRNA. **a**, *Renilla* luciferase reporter mRNAs with three *let-7*-binding sites (RL 3XB WT; top), three mutated sites (RL 3XB mt; middle) and lacking a 3' UTR (RL; bottom). **b–d**, Pre-miRNA turnover using N2 lysate in the presence of the indicated mRNAs. Substrates were radiolabelled pre-*let-7* (**b**), unlabelled pre-*let-7* analysed by northern blotting (**c**) and radiolabelled pre-*mir-237* (**d**). Right arrowheads: mature *mir-237* (23 nt) and *let-7* (22 nt). Lanes in **c** are all from a single autoradiograph. **e**, Top: pre-*let-7* turnover assay as in **b**, using *gfp::ago* lysates. Middle: mature *let-7* co-immunoprecipitated with an α -GFP antibody. Bottom: post-immunoprecipitation supernatant (separate gel).

because both *let-7* levels and activity are increased in *xrn-2(RNAi)* animals (Fig. 1, Supplementary Fig. 3 and Supplementary Table 1).

Although our data support the notion that miRNAs can be released from miRISC, this result is unexpected because human siRNA–AGO complexes are highly stable²⁶. To test the stability of miRNA–AGO complexes, we immunoprecipitated GFP/AGO from larval lysate either immediately or after incubation at 25 °C. Levels of *let-7* decreased in immunoprecipitate obtained after the incubation step relative to the immunoprecipitate obtained before incubation (Fig. 4a, compare lane 1 and 2). Because the levels and integrity of GFP/AGO were unaltered, this finding supports the disassembly of the miRNA–AGO complex.

To demonstrate directly the release of AGO-bound miRNA, we immunoprecipitated GFP/AGO and incubated the protein, while bound to beads, with assay buffer with or without added KCl, or micrococcal nuclease/EGTA-treated lysates from worms exposed to mock RNAi or *xrn-2(RNAi)*. After recovery, RNA from the beads and the supernatants was probed by northern blotting for the presence of endogenous *let-7*. As expected, the addition of neither buffer nor a high concentration of salt diminished the level of AGO-bound *let-7* relative to the control, and no signal was detected in the supernatants (Fig. 4b; compare lanes 1–3). By contrast, incubation with wild-type larval lysate resulted in a strong loss of *let-7* signal from the beads and no signal in the supernatant, which is consistent with the removal of *let-7* from the ALG-1/2 complexes and its subsequent degradation (Fig. 4b, lane 4).

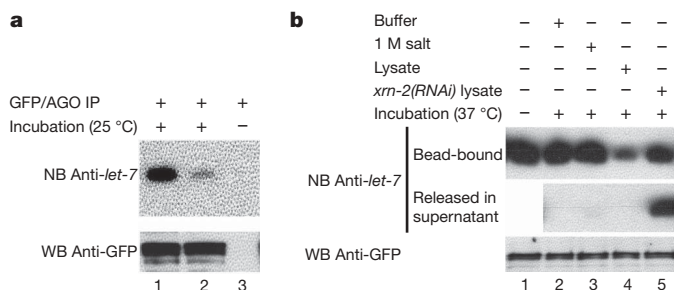


Figure 4 | Release of miRNA from miRISC. **a**, GFP/AGO immunoprecipitation from lysate directly or after incubation for 15 min. Half of each immunoprecipitate was subjected to northern analysis, the other half to anti-GFP western blotting. Lane 3, no antibody. **b**, Immunoprecipitated GFP/AGO proteins processed immediately (lane 1), or after incubation for 15 min with assay buffer (AB, lane 2), AB + salt (lane 3), N2 (lane 4) or *xrn-2(RNAi)* (lane 5) lysate. RNA from supernatant and from half of the bead-bound reaction were subjected to northern analysis (top and middle, respectively). The other half of the bead-bound material was subjected to anti-GFP western blotting (bottom).

When *xrn-2(RNAi)* lysate was used to induce release, more mature miRNA remained bound to GFP/AGO (Fig. 4b; compare lanes 4 and 5), supporting a facilitating role of XRN-2 in the release process. Substantial amounts of *let-7* could be recovered in the supernatant under these conditions, confirming that efficient degradation of released miRNA depends on XRN-2.

We have shown here that XRN-2 mediates miRNA turnover *in vivo* and *in vitro*, and that it can modulate the activity of miRNAs *in vivo*. Thus, miRNA degradation contributes to miRNA homeostasis, helping to prevent detrimental overexpression of miRNAs¹. The fact that miRNAs can stabilize their cognate miRNAs *in vitro* suggests a co-ordination of miRNA and target levels (Supplementary Fig. 1), permitting miRISC reprogramming when target abundance is low. When miRNA silencing is prevented or reversed^{27,28}, increased degradation of the concerned miRNA(s) might further enhance desilencing by preventing miRISC from target rebinding.

Although we find that XRN-2 is important in miRNA turnover *in vitro*, not all miRNAs accumulate efficiently on *xrn-2* depletion *in vivo*. It will be interesting to determine whether this reflects genuine substrate specificity or incomplete depletion of XRN-2 at times or in tissues of greatest expression of the less-affected miRNAs. It also seems likely that specialized regulatory proteins, and possibly additional ribonucleases, remain to be discovered that will guide differential turnover of individual miRNAs, for example to modulate expression profiles of specific miRNAs during developmental transitions.

METHODS SUMMARY

To prepare lysates, worms were suspended in extraction buffer (10 mM HEPES pH 7.4, 2 mM dithiothreitol (DTT), 0.1% Triton X-100, 50 mM KCl, 0.5 mM phenylmethylsulfonyl fluoride (PMSF), 10% glycerol), ground in liquid N₂ and cleared by centrifugation. RNA substrates were incubated with lysate in 1 × assay buffer (AB; 10 mM HEPES pH 7.4, 2 mM DTT, 5 mM MgCl₂, 100 mM KCl, 2 mM ATP).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 April; accepted 27 July 2009.
Published online 6 September 2009.

- Chang, T. C. & Mendell, J. T. microRNAs in vertebrate physiology and human disease. *Annu. Rev. Genomics Hum. Genet.* **8**, 215–239 (2007).
- Ding, X. C., Weiler, J. & Großhans, H. Regulating the regulators: mechanisms controlling the maturation of microRNAs. *Trends Biotechnol.* **27**, 27–36 (2009).
- Esquela-Kerscher, A. & Slack, F. J. Oncomirs—microRNAs with a role in cancer. *Nature Rev. Cancer* **6**, 259–269 (2006).

- Ramachandran, V. & Chen, X. Degradation of microRNAs by a family of exoribonucleases in *Arabidopsis*. *Science* **321**, 1490–1492 (2008).
- Büssing, I., Slack, F. J. & Großhans, H. *let-7* microRNAs in development, stem cells and cancer. *Trends Mol. Med.* **14**, 400–409 (2008).
- Reinhart, B. J. et al. The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901–906 (2000).
- Vella, M. C., Choi, E. Y., Lin, S. Y., Reinert, K. & Slack, F. J. The *C. elegans* microRNA *let-7* binds to imperfect *let-7* complementary sites from the *lin-41* 3' UTR. *Genes Dev.* **18**, 132–137 (2004).
- Bagga, S. et al. Regulation by *let-7* and *lin-4* miRNAs results in target mRNA degradation. *Cell* **122**, 553–563 (2005).
- Abbott, A. L. et al. The *let-7* microRNA family members *mir-48*, *mir-84*, and *mir-241* function together to regulate developmental timing in *Caenorhabditis elegans*. *Dev. Cell* **9**, 403–414 (2005).
- Kennedy, S., Wang, D. & Ruvkun, G. A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature* **427**, 645–649 (2004).
- Chernyak, I., Whipple, J. M., Kotelawala, L., Grayhack, E. J. & Phizicky, E. M. Degradation of several hypomodified mature tRNA species in *Saccharomyces cerevisiae* is mediated by Met22 and the 5'–3' exonucleases Rat1 and Xrn1. *Genes Dev.* **22**, 1369–1380 (2008).
- Weidhaas, J. B. et al. MicroRNAs as potential agents to alter resistance to cytotoxic anticancer therapy. *Cancer Res.* **67**, 11111–11116 (2007).
- Gy, I. et al. *Arabidopsis* FIERY1, XRN2, and XRN3 are endogenous RNA silencing suppressors. *Plant Cell* **19**, 3451–3461 (2007).
- Lee, R. C. & Ambros, V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**, 862–864 (2001).
- Ketting, R. F. et al. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**, 2654–2659 (2001).
- Grishok, A. et al. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**, 23–34 (2001).
- Großhans, H., Johnson, T., Reinert, K. L., Gerstein, M. & Slack, F. J. The temporal patterning microRNA *let-7* regulates several transcription factors at the larval to adult transition in *C. elegans*. *Dev. Cell* **8**, 321–330 (2005).
- Slack, F. J. et al. The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the *let-7* regulatory RNA and the LIN-29 transcription factor. *Mol. Cell* **5**, 659–669 (2000).
- Ding, X. C. & Großhans, H. Repression of *C. elegans* microRNA targets at the initiation level of translation requires GW182 proteins. *EMBO J.* **28**, 213–222 (2009).
- Stevens, A. & Poole, T. L. 5'-exonuclease-2 of *Saccharomyces cerevisiae*. Purification and features of ribonuclease activity with comparison to 5'-exonuclease-1. *J. Biol. Chem.* **270**, 16063–16069 (1995).
- Stevens, A. & Maupin, M. K. A 5'–3' exoribonuclease of human placental nuclei: purification and substrate specificity. *Nucleic Acids Res.* **15**, 695–708 (1987).
- Hutvagner, G. et al. A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science* **293**, 834–838 (2001).
- Pillai, R. S. et al. Inhibition of translational initiation by *let-7* microRNA in human cells. *Science* **309**, 1573–1576 (2005).
- Hutvagner, G., Simard, M. J., Mello, C. C. & Zamore, P. D. Sequence-specific inhibition of small RNA function. *PLoS Biol.* **2**, E98 (2004).
- Wang, Y., Sheng, G., Juranek, S., Tuschl, T. & Patel, D. J. Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456**, 209–213 (2008).
- Martinez, J. & Tuschl, T. RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev.* **18**, 975–980 (2004).
- Bhattacharyya, S. N., Habermacher, R., Martiny-Bar, U., Closs, E. I. & Filipowicz, W. Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* **125**, 1111–1124 (2006).
- Kedde, M. et al. RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* **131**, 1273–1286 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Bühler and W. Filipowicz for critical comments on the manuscript; W. Filipowicz for plasmids; C. Mello and F. Slack for *C. elegans* strains; and A. Esquela-Kerscher for sharing Starfire probes for pre-miRNA detection. S.C. was supported by Marie Curie and EMBO long-term postdoctoral fellowships.

Author Contributions S.C. and H.G. designed the research. S.C. designed and performed the experiments. S.C. and H.G. analysed the experimental results and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.G. (helge.grosshans@fmi.ch).

METHODS

Worm strains, *let-7(n2853)* suppression and RNAi. The wild-type strain was *C. elegans* var. Bristol strain N2. The other two strains were *let-7(n2853)* (ref. 6) and *gfp::alg-1;gfp::alg-2* (ref. 24). Suppressors of *let-7(n2853)* were identified by RNAi by feeding growing worms on RNAi plates at 25 °C as described^{17,29}.

RNA isolation, northern blotting and RT-qPCR. Total RNA was isolated from staged L4 worms, unless indicated otherwise, that were frozen in liquid nitrogen and ground in a mortar, using the Trizol (Invitrogen) method in accordance with the manufacturer's instruction. Northern blotting of endogenous RNA was performed as described³⁰. 5'-labelled (using T4 polynucleotide kinase (PNK) and [γ -³²P]ATP) DNA oligos were used as probes except for those in Supplementary Fig. 5, in which STARFIRE probes were used. The hybridization for *let-7* miRNA was conducted at an elevated temperature of 40 °C to minimize the binding of the probe to *let-7* sisters. For northern analysis of *in vitro* turnover assay products, the lysates were pretreated with micrococcal nuclease (NEB; 0.5–1.0 μ l per 100 μ g of lysate) for 10 min at 37 °C followed by the addition of EGTA to a final concentration of 7.5 mM. Pretreatment with micrococcal nuclease was done to digest all endogenous RNAs from the lysates, ruling out the possibility of detection of endogenous RNA. Excess EGTA was used to terminate the micrococcal nuclease treatment. '–EGTA' lysate served as a positive control for micrococcal nuclease activity that removed all the RNA, including exogenous RNA, resulting in no signal. After incubation of RNAs in the lysates, the samples were extracted with phenol/chloroform and precipitated with alcohol. The recovered samples were subjected to northern probing using the conditions stated above. RT-qPCR was performed as described¹⁹.

Cloning and expression of recombinant XRN-2. *xrn-2* complementary DNA was amplified from total RNA by RT-PCR, and cloned in a TOPO TA vector (Invitrogen). The sequence-confirmed correct ORF was subcloned in pGEX 4T-1 (GE Healthcare) and expressed in *Escherichia coli* as a GST fusion protein. The recombinant protein was extracted with detergent from inclusion bodies and then resolved by SDS-PAGE. After KCl staining³¹, the band of pure recombinant protein was excised and the protein was eluted with buffer PEB (0.05 M Tris-HCl pH 8.0, 0.2 M NaCl, 0.1 mM EDTA, 5 mM DTT, 0.2% SDS) overnight at 37 °C. The eluate was stored in aliquots at –80 °C. Before use, the recombinant protein was refolded by diluting 1:20 in TETN 250 buffer containing 0.1% Triton X 100 (ref. 32), and incubated for 2 h at 4 °C. Finally the protein was concentrated with Microcon-100 (Millipore) and its concentration was estimated with a Bradford assay (Bio-Rad). The use of gel-purified and renatured GST-XRN-2 (about 140 kDa) essentially eliminated any possibility of contamination by 5'→3' exoribonuclease activity of bacterial origin because proteins larger than 100 kDa are rare in *E. coli*. Moreover, bacterial cells are thought to be largely devoid of endogenous 5'→3' exoribonucleases, with the only known bacterial activity residing in a 66-kDa protein, the *B. subtilis* endoribonuclease/exoribonuclease ribonuclease J1 (ref. 33).

Preparation of RNA substrates. Pre-*let-7*/pre-*mir-237* or mature *let-7* RNA were prepared essentially by following the methods described elsewhere³⁴. In brief, a chimaeric RNA containing in its 5' portion a hammerhead ribozyme followed by the pre-*let-7*/mature *let-7* sequence was transcribed from DNA cassettes with a T7 MAXIScript kit (Ambion) in the presence of [α -³²P]UTP or unlabelled UTP. The DNA cassettes were prepared by the annealing of appropriate forward and reverse primers (see the oligo sequence section) followed by Klenow fill-in reactions. Double-stranded DNAs of appropriate lengths were gel-purified and amplified by PCR (except pre-*mir-237* and mature *let-7*, which were directly used as transcription substrate after gel purification) with appropriate flanking primers. Gel-purified PCR products were used as the templates for *in vitro* transcription reactions. Moreover, before use, the PCR products were cloned and their sequences were confirmed. Self-processing of the ribozyme-containing transcripts occurred during the course of transcription reaction. The resulting pre-*let-7*/mature *let-7*, which contained 5' hydroxyl groups, were size-purified with 8–10% PAGE in the presence of 7 M urea. After recovery, RNAs were 5' phosphorylated with T4 PNK and ATP. Before use the pre-*let-7* RNA was subjected to refolding as described³⁴.

5' labelling of synthetic mature miRNAs and tRNA (after dephosphorylation; yeast tRNA^{Phe}; Sigma) was performed with PNK and [γ -³²P]ATP. 3' labelling and blocking of synthetic mature miRNAs were performed with T4 RNA ligase and [5'-³²P]pCp, in accordance with the manufacturer's instructions (Ambion). All RNAs were gel purified.

The 5'-7-methyl-G-capped RL reporter mRNAs²³ (*Renilla* luciferase with an artificial 3' UTR harbouring 3X bulged *let-7*/miR-237 complementary sites (RL 3XB WT), *Renilla* luciferase followed by mutated 3X bulged *let-7* complementary sites (RL 3XB mt), and *Renilla* luciferase without 3' UTR (RL)) were prepared by *in vitro* run-off transcription of appropriately digested plasmids using standard reagents from a T7 MEGAScript kit and cap analogue m⁷G(5')ppp(5')G from

Ambion, in accordance with the manufacturer's instructions. After extraction with phenol/chloroform and precipitation with alcohol, the RNAs were polyadenylated with *E. coli* PolyA polymerase (Stratagene) and ATP.

The construct for RL mRNA with an artificial 3' UTR harbouring 3X bulged *mir-237* target sites were prepared by swapping the bulged *let-7* complementary sites in the pRL 3XB WT vector²³ with bulged *mir-237* complementary sites.

Preparation of worm lysate. Staged L4 worms grown on plates were harvested with M9 buffer (22 mM KH₂PO₄, 22 mM Na₂HPO₄, 85 mM NaCl, 1 mM MgSO₄) and washed three times with the same buffer. The worm pellet was then resuspended in extraction buffer (10 mM HEPES pH 7.4, 2 mM DTT, 0.1% Triton X-100, 50 mM KCl, 0.5 mM PMSF, 10% glycerol) and ground in liquid N₂. After thawing, the sample was centrifuged at 14,000g or more for 15–20 min; the clear supernatant was collected and designated as cleared worm lysate.

***In vitro* turnover assay.** Labelled RNAs (pre-*let-7*/pre-*mir-237* and mature *let-7*, about 1 and 2 fmol, respectively) were incubated for 15 min with cleared worm lysate (2–20 μ g) in 1 \times assay buffer (AB; 10 mM HEPES pH 7.4, 2 mM DTT, 5 mM MgCl₂, 100 mM KCl, 2 mM ATP) in a volume of 10 μ l at 25 or 37 °C. The reactions were stopped by addition of 1 volume of formamide gel loading buffer (95% formamide, 0.2% SDS, 1 mM EDTA, 0.04% xylene cyanol, 0.04% bromophenol blue) followed by heating at 65 °C for 5 min. Equal volumes of the samples were then subjected to 8–12% PAGE in the presence of 7 M urea, followed by gel drying and autoradiography or phosphorimaging.

The target-mRNA-mediated miRNA stabilization assays were performed with 1 fmol of radiolabelled substrate and 20 fmol of the concerned target mRNA in volumes of 20 μ l.

For add-back assays, the *xrn-2(RNAi)* lysate was preincubated with the recombinant protein (about 7 ng per reaction) on ice for 30 min to achieve reconstitution, and then used for the assay.

For native gel analysis of pre-*let-7* assay products, reactions were performed as above except that after incubation the samples were subjected to treatment with proteinase K (PK) as described³⁵ at 20–23 °C for 30 min, and then resolved in a 15% native polyacrylamide gel at 4 °C, using a native gel loading buffer³⁵.

Guide-passenger duplex with a 5'-³²P-labelled guide strand was prepared using methods described previously³⁵; 10,000 c.p.m. of the native gel purified substrate was used in each reaction. After incubation with lysate, analysis of the sample was performed under native conditions as for pre-*let-7* (above).

Coupled pre-*let-7* processing and Ago immunoprecipitation. Pre-*let-7* assay was performed as described above in the absence or presence of target mRNA, using a lysate obtained from a strain in which both the *C. elegans* miRISC Argonaute proteins ALG-1 and ALG-2 were tagged with GFP ('GFP/AGO'; ref. 24). After incubation for 15 min at 37 °C, the reaction volumes were increased to 200 μ l with 1 \times AB and subjected to immunoprecipitation at 4 °C for 2 h with an anti-GFP antibody (anti-GFP mouse IgG; monoclonal antibody, catalogue no. 11 814 460 001; Roche) and Protein A-Sepharose CL-4B (GE Healthcare). The recovered Sepharose beads were suspended in formamide gel loading buffer, heated at 65 °C for 5 min and centrifuged briefly; the supernatants were subjected to urea-PAGE analysis. The post-immunoprecipitate supernatants were also recovered through phenol/chloroform extraction and alcohol precipitation, and subjected to urea-PAGE analysis.

Thin-layer chromatography. Mature miRNA turnover reactions using [α -³²P]UTP-labelled miRNA were stopped by the addition of SDS to 1% concentration and EDTA to 10 mM. Aliquots of 1 μ l were spotted onto prewashed PEI-cellulose plates (Macherey-Nagel) and developed sequentially with 0.5 M LiCl and 1 M formic acid³⁶. Unlabelled uridine 5'-monophosphate and uridine 5'-diphosphate were also separated on the same plate and detected through fluorescence quenching.

miRNA release assay. Immunoprecipitation of GFP-tagged ALG-1/ALG-2 complexes was performed essentially by following the methods described elsewhere³⁷ with the aforementioned anti-GFP antibody. The bead-bound immunoprecipitates (derivative of 400 μ g of lysate protein per reaction) were incubated with 1 \times AB, 1 \times AB plus KCl (to a final concentration of about 1.0 M) or 100 μ g of micrococcal nuclease/EGTA-treated empty vector and *xrn-2(RNAi)* lysate, at 37 °C for 15 min. After further recovery the beads were split into two halves. RNA was extracted from one half for northern analysis; the other half was boiled in SDS sample buffer and subjected to SDS-PAGE and western blotting with a rabbit polyclonal anti-GFP antibody (ab6556-25; Abcam), to confirm equal binding of AGO to the beads and integrity of the proteins. RNA was also extracted from the above supernatant fractions (not split) and subjected to northern analysis to detect miRNAs released in the supernatant.

In the parallel approach, immunoprecipitation was performed from lysate before and after incubation at the worm's physiological temperature (25 °C) for 15 min, and the immunoprecipitate was subjected to both northern and western probing as mentioned above.

Oligos (5'-3'). Northern: *let-7*(WT), 5'-AAC TAT ACA ACC TAC TAC CTC A-3'; *let-7*(n2853), 5'-AAC TAT ACA ACC TAC TAT CTC A-3'; *mir-77*, 5'-TG G ACA GCT ATG GCC TGA TGA A-3'; *mir-85*, 5'-GCA CGA CTT TTC AAA TAC TTT GTA-3'; *pre mir-60*, 5'-CT TGA ACT AGA AAA TGT GCA TAA TA TCA CGT ACT TTG TCA TG-3'; *tRNA^{Gly}*, 5'-GCTTGGAAGGCATCCATG CTGACCATT-3'. STARFIRE probes (IDT) against the following miRNAs were used: *let-7*, *mir-237*, *lin-4*, *mir-34*, *mir-240*, *mir-75*, *mir-245*, *mir-234*, *mir-79*, *mir-84*, *mir-48*, *mir-85*.

qPCR: primary *let-7*, 5'-TCCTAGAACACATCTCCCTTTGA-3' (forward) and 5'-CGCAGCTTCGAAGAGTTCTG-3' (reverse); primary *mir-77*, 5'-CATT GTTCGTTTCGCTTTCA-3' (forward) and 5'-CCAATAACTGATTCAACATT CCAA-3' (reverse); *daf-12* mRNA, 5'-GAT CCT CCG ATG AAC GAA AA-3' (forward) and 5'-CTC TTC GGC TTC ACC AGA AC-3' (reverse); *lin-41* mRNA, 5'-GGA TTG TTC GAC ACC AAC G-3' (forward) and 5'-ACC ATG ATG TCA AAC TGC TGT C-3' (reverse); *xrn-2* mRNA, 5'-GATCCCGAGTACCCA CAAGA-3' (forward) and 5'-CCACCACCACCTCTCACATA-3' (reverse).

Cloning: *xrn-2* cDNA, 5'-GAAA GAATTC ATG GGA GTT CCC GCA TTC TTC AG-3' (forward primer) and 5'-GAAA GCGGCCGC GAT TAT CTC CAT GAT GAA TTT CCG TG-3' (reverse primer). 3X *mir-237* target cassette construction: template sequence, 5'-GGGG tctaga AGC TGT TCG AGA ATT TTGAA CTC AGG GA ctcggagc AGC TGT TCG AGA ATT TTGAA CTC AGG GA ctcggagc AGC TGT TCG AGA ATT TTGAA CTC AGG GA ctcggagc AAAG-3'. Primers for PCR amplification of 3X *mir-237* target cassette: 5'-GGG GTC TAG AAG CTG TTC GAG AAT TTT G-3' (forward primer) and 5'-CTT TGC GGC CGC TCC CTG AGT TCA AAA TTC-3' (reverse primer).

Preparation of templates for *in vitro* transcription: mature *let-7* cassette, 5'-G TAA TAC GAC TCA CTA TAG GGAGA CTA CTA CCT CAC TGA TGA GTC CGT GAG GAC GAA ACG GTA CCC GGT ACC GTC TGA GGT AGT AGG-3' (forward primer (T7 promoter, HH ribozyme, first 12 *let-7* nucleotides)) and 5'-AAC TAT ACA ACC TAC TAC CTC A GAC GGT ACC GGG-3' (reverse primer (mature *let-7* complementary sequence, 12 nucleotides; complementary region to HH ribozyme)). *Pre-let-7* cassette: 5'-G TAA TAC GAC TCA CTA TAG GGAGA CTA CTA CCT CAC TGA TGA GTC CGT GAG GAC GAA ACG GTA CCC GGT ACC GTC TGA GGT AGT AGG-3' (forward primer (T7 promoter, HH ribozyme, first 12 *let-7* nucleotides)) and 5'-GGT AAG GTA GAA AAT TGC ATA GTT CAC CGG TGG TAA TAT TCC AAA CTA TAC AAC CTA CTA CCT CA GAC GGT ACC GGG-3' (reverse primer (pre-*let-7* complementary sequence, 12 nucleotides; complementary region to HH ribozyme)).

Primers for PCR amplification of mature *let-7* cassette: 5'-GAATTC TAA TAC GAC TCA CTA TAG G-3' (forward T7 promoter primer) and 5'-AAC TAT ACA ACC TAC TAC CTC A-3' (*let-7* guide reverse primer). Primers for PCR amplification of pre-*let-7* cassette: 5'-GAATTC TAA TAC GAC TCA CTA TAG G-3' (forward T7 promoter primer) and 5'-GGT AAG GTA GAA AAT TGC ATA G-3' (*let-7* passenger reverse primer). Pre-*mir-237* cassette: 5'-G TAA TAC GAC TCA CTA TAG GGAGA GAA TTC TCA GGG A C TGA TGA GTC CGT GAG GAC GAA ACG GTA CCC GGT ACC GTC TCC CTG AGA ATT C-3' (forward primer (T7 promoter, HH ribozyme, first 13 *mir-237* nucleotides)) and 5'-GGT CCT TGA CAA AAC TCG ACA GCT TGA ACA CTT TGA AGC TGT TCG AGA ATT CTC AGG GAGAC GGT ACC GGG-3' (reverse primer (pre-*mir-237* complementary sequence, 12 nucleotides; complementary region to HH ribozyme)).

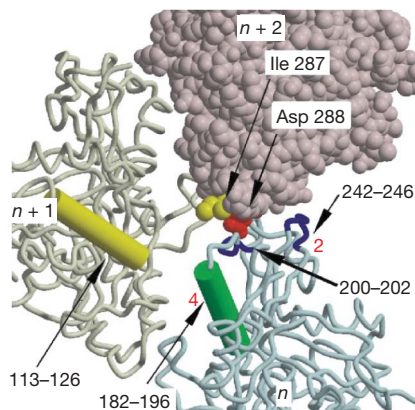
29. Ding, X. C., Slack, F. J. & Großhans, H. The *let-7* microRNA interfaces extensively with the translation machinery to regulate cell differentiation. *Cell Cycle* **7**, 3083–3090 (2008).
30. Pall, G. S. & Hamilton, A. J. Improved northern blot method for enhanced detection of small RNA. *Nature Protocols* **3**, 1077–1084 (2008).
31. Hager, D. A. & Burgess, R. R. Elution of proteins from sodium dodecyl sulfate-polyacrylamide gels, removal of sodium dodecyl sulfate, and renaturation of enzymatic activity: results with sigma subunit of *Escherichia coli* RNA polymerase, wheat germ DNA topoisomerase, and other enzymes. *Anal. Biochem.* **109**, 76–86 (1980).
32. Chatterjee, S. *et al.* An RNA-binding respiratory component mediates import of type II tRNAs into *Leishmania* mitochondria. *J. Biol. Chem.* **281**, 25270–25277 (2006).
33. Mathy, N. *et al.* 5'-to-3' exonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5' stability of mRNA. *Cell* **129**, 681–692 (2007).
34. Kolb, F. A. *et al.* Human dicer: purification, properties, and interaction with PAZ PIWI domain proteins. *Methods Enzymol.* **392**, 316–336 (2005).
35. Matranga, C., Tomari, Y., Shin, C., Bartel, D. P. & Zamore, P. D. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123**, 607–620 (2005).
36. Dziembowski, A., Lorentzen, E., Conti, E. & Séraphin, B. A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nature Struct. Mol. Biol.* **14**, 15–22 (2007).
37. Lee, M. H. & Schedl, T. Identification of *in vivo* mRNA targets of GLD-1, a maxi-KH motif containing protein required for *C. elegans* germ cell development. *Genes Dev.* **15**, 2408–2420 (2001).

CORRIGENDUM

doi:10.1038/nature08440

The nature of the globular- to fibrous-actin transitionToshiro Oda, Mitsusada Iwasa, Tomoki Aihara, Yuichiro Maéda
& Akihiro Narita*Nature* 457, 441–445 (2009)

In this Article, Figure 4a was incorrect. The correct figure is shown below. The results and conclusions of the Article are not affected.



CORRIGENDUM

doi:10.1038/nature08464

Reptilian heart development and the molecular basis of cardiac chamber evolution

Kazuko Koshiba-Takeuchi, Alessandro D. Mori, Bogac L. Kaynak,
Judith Cebra-Thomas, Tatyana Sukonnik, Romain O. Georges,
Stephany Latham, Laurel Beck, R. Mark Henkelman, Brian L. Black,
Eric N. Olson, Juli Wade, Jun K. Takeuchi, Mona Nemer,
Scott F. Gilbert & Benoit G. Bruneau

Nature 461, 95–98 (2009)

In this Letter, author Laurel Beck was incorrectly listed as Laural Beck.

ERRATUM

doi:10.1038/nature08491

Structural basis for leucine-rich nuclear export signal recognition by CRM1

Xiuhua Dong, Anindita Biswas, Katherine E. Süel, Laurie K. Jackson, Rita Martinez, Hongmei Gu & Yuh Min Chook

Nature 458, 1136–1141 (2009)

In the first paragraph of this Article, Leu was incorrectly listed as Lys. The correct sentence is: “Several mutagenesis and computational studies have led to a consensus sequence ϕ -X(2–3)- ϕ -X(2–3)- ϕ -X- ϕ (in which ϕ is Leu, Val, Ile, Phe or Met, X is any amino acid, and the numbers in parentheses denote the number of repeats) that matches most known LR-NESs, but it is so broad that it is found in most helix-containing proteins^{17,18,19,20,21}.”

A kiss isn't just a kiss

Mouth-to-mouth resuscitation.

Steve Carper

On its slowest day, Hong Kong International Airport fed tens of thousands of bodies into the jets bearing the insignia of 60 different airlines. Today, the bodies jostled one another in lines that boiled and roiled and dissolved and reformed as loved ones gave one another parting kisses. Then kissed friends, children, strangers, staff, crew, baggage handlers, taxi drivers and police, who kept threatening them with their clubs. The flu had arrived and they were the first wave out of the infected area.

Flu.

In Chinese and English the word jumped at them, from newspaper headlines to the captioning of CNN International and Sky-Net to the announcement tickers above every bank of screens. They were heading across the world: direct flights to almost every continent. Thousands would greet them at every airport, braving the clubs of their country's police and armed forces for the wonderful, life-granting prize of all prizes, the other word blazoned across every means of communication in the airport.

Cure.

The first signs of the avian flu strain to be labelled H5LB7 appeared in villages in northern Vietnam in early February. Authorities quickly set up roadblocks and sent in teams to destroy the infected chickens, but rumours of an extraordinarily high percentage of deaths filtered out. It was later revealed that dead birds and dead humans alike were spirited across the border to the National Key Laboratory of Veterinary Biotechnology in Harbin.

As the flu spread across Vietnam into Laos and Guangxi and Guangdong provinces, virologists adapted a technique that had earlier proved useful against the H5N1 strain. Unlocking the virus's genome, they cloned segments of its DNA and proceeded to mutate genes and reintroduce them into the genome. By July they had a stable strain of the virus that acted as a viable vaccine.

It was almost too late. Shenzhen Airlines was no longer flying out of Nanning in the affected provinces. Xiamen Airlines, Shanghai Airlines, Sichuan Airlines and both China Eastern and China Southern Airlines had been forced to shut down all but a bare few emergency flights to and from surrounding countries.

"The vaccine works," reported Dr Cao Xiping. "The new virus displays both the haemagglutinin and neuraminidase antigens. We added to those genes a half-dozen

others known for their ability to grow inside eggs as a base medium, which we can then refine to capture the vaccine."

"Eggs?" asked the agriculture minister. "How long to grow a billion doses inside eggs?"

Dr Cao gave the number. Too long. Too expensive. Too difficult. Everything was against it. The minister tried to wipe the look of scared horror off his face sufficiently to look angry and threatening, with minimal success.

"We do have another idea," said Dr Zhu Yi. "Naturally, it has never been tested. But it does have the advantage of the quickest and cheapest wide-scale dissemination of the vaccine practicable."

"Let's hear it," the minister said.

He didn't like it. He was against it. He went into a tirade that would have intimidated anyone not as bone-tired and single-minded as the virologists had become after weeks of unending death. The meeting went on for hours. At midnight the minister capitulated. He did not look happy. His expression did not change even after Dr Zhu gave him a long and somewhat sloppy kiss. He wiped his mouth with a napkin left over from the dinner and hurried home to kiss his wife and children.

None of them contracted the flu.

Technicians at the National Key Laboratory started incubating eggs, while virologists mutated more genes on the already mutated virus. Doses of the vaccine were refined at almost the same rate and number as the list of airports, seaports and border checkpoints being sealed off because of the spreading infection.

Meetings at the Ministry of Agriculture grew in similar hypertrophic fashion, as representatives of the ministries of health, science and technology, communications, foreign affairs, state security and national defence joined around ever-larger tables. Their decision went out through all state-run media. Volunteers would gather at the Hong Kong International Airport in one week's time. All travel expenses would be paid. So would their burial expenses if anything went wrong, although that part was whispered to them as they gathered.

The news exploded onto the world to be greeted with an array of emotions as vast as all humanity. The Chinese had perfected a cure for the deadliest strain of flu ever encountered. There was no time to inoculate everyone. A faster mode of transmission was necessary. And the fastest mode of transmission known to mankind was ... mankind.



Dr Zhu and Dr Cao had modified the virus/vaccine to be transmissible by human-to-human contact. It lived — thrived! — in saliva. Kissing was an exemplary use of the human vector. People met people faster than any other form of contact, short of spraying from aeroplanes. There were technical reasons why that wasn't feasible, but they were lost under the sounds of the explosion as the world learned that to save itself it had to — as the British tabloids headlined over and over — *snog*.

The first wave funnelled through the newly reopened Hong Kong International Airport on Chek Lap Kok, kissing everyone within reach. They fanned out through the cities of dozens of nations, an exponential force, disinfecting dozens in each. Men kissed men, women kissed women, all kissed children and babies, and those who fought the contact were, it is accurate to say, sometimes kissed against their will. Few of those who successfully resisted died, but that was because by then so few successful carriers surrounded them.

The H5LB7 flu killed several million people in south Asia and smaller numbers elsewhere. About a billion people who might otherwise have died did not.

Some of them actually looked forward to the next year's flu season. ■

Steve Carper's collection of fantasy and science-fiction short stories is *Tyrannosaur Faire*. He is editing the next edition of the *SFWA Handbook* and is a columnist for the *SFWA Bulletin*. Join the discussion of *Futures in Nature* at <http://tinyurl.com/kkh3kt>

NEWS

Sponsor a scientist, online

In these lean times, research institutions need to be creative in the quest for extra funding. The latest brainstorm is online bidding to encourage donor support for graduate research fellowships.

'Sponsor-a-Scientist' is an item on eBay, the online auction site, as of 4 September. The Sbarro Health Research Organization (SHRO), a non-profit collective based at Temple University in Philadelphia, Pennsylvania, and the University of Siena in Italy, hopes it will yield at least one fellowship, to last a year, for graduate training. A best-case scenario would be five fellowships.

Bids start at US\$25,000. The winner will be entitled to name the fellowship (in their name or to honour someone else), meet the sponsored researcher and tour the lab. Raffaella Cimina, director of development at the SHRO, which specializes in research into cancer, diabetes and cardiovascular disease, conceived the idea. At the time *Nature* went to press, there were no bids, but Cimina says they plan to renew the bid every ten days in the hope that word will spread and bids will materialize before the auction's target end date, tentatively planned for early 2010.

Some 120 graduate students and first-year postdocs are being trained at the SHRO to investigate the genetics of lung, breast and prostate cancer, hepatitis and cardiovascular disease, as well as stem-cell research and developmental biology. But despite success in securing grants from the US National Institutes of Health and state and private donors, raising the money needed to train graduates and postdocs

is difficult. "Given the tightness of money for training new scientists and the pervasiveness of online portals and social networking, using the



Internet to jumpstart a scientific career seemed a natural next step," says SHRO president and chief executive Antonio Giordano. He says that SHRO wants to take advantage of sites such as eBay to reach new donors and engage the public in biomedical research. He also hopes to build connections between donors

and researchers.

Steven Perrin, president of the non-profit ALS Technology Development Institute (ALS TDI) in Boston, Massachusetts, suggests that the SHRO's eBay approach could, if carefully implemented, ensure the long-term support needed for several years of student training. Giordano guarantees that the recipient of any fellowship will be covered for their entire training period by matching funds from existing donations.

"We all need to be creative to keep non-profits productive and moving forward," says Perrin, whose organization focuses on amyotrophic lateral sclerosis, a progressive neurodegenerative disease. The ALS TDI has already used social-networking sites, such as Facebook, to help connect those interested in research fund-raising for a particular disease.

If the eBay approach works, Giordano says the SHRO will expand the idea — and he expects that other institutions will follow its lead. ■

Virginia Gewin

Sponsor-A-Scientist is at
<http://tinyurl.com/n5l2q2>

Lab trips foster collegiality



Every summer the members of our lab spend a weekend at a state park, and our most recent annual outing was to Niagara Falls, where we camped, hiked and enjoyed the scenery. I was captivated by the thunderous roar and sprays of mist emanating from the falls. I always look forward to the lab outing because it is guaranteed to be fun — as it was this summer.

But these outings are more than just fun excursions. They also build collegiality among our lab members. This is important because cooperation and effective

communication can help the lab to run more smoothly and more efficiently. During summer outings, lab members work together to set up tents, build a campfire and cook meals. We exchange confidences during long hikes. The value in building these interactions is reflected in the operations of the lab, as co-workers must similarly depend on each other for reagents, interpreting results and troubleshooting.

In a different lab, I once watched two graduate students troubleshooting related molecular-biology

protocols. One eventually discovered that a labile buffer had gone bad. However, he selfishly opted to keep the knowledge to himself, leaving his fellow student to flounder for weeks. A lack of collegiality can have a serious impact on lab members. Consequently, in the near future when I look for another job, collegiality is a factor I will consider. I have witnessed at first hand its value in the workplace. ■

Bryan Venters is a postdoctoral fellow at the Center for Eukaryotic Gene Regulation at Pennsylvania State University.

IN BRIEF

Biotech goes to school

A bill passed by the California state legislature should make it easier for the California biotech industry to recruit technical staff from within the state. The bill, which still needs the governor's signature, requires the state education department to incorporate stem-cell and biotech curricula into the public schools, with help from industry and the California Institute for Regenerative Medicine (CIRM), the state's stem-cell agency. A state biotech trade association warned in a 2008 report of a scarcity of trained lab professionals. "The pipeline starts in high school, and having the curriculum at this level is really important," says CIRM spokesman Don Gibbons.

Teaching sequestration

A new training programme for prospective workers in the carbon capture and storage industry will target geoscientists and engineers currently working in the oil industry as well as academic scientists. Under development by the Alliance for Sequestration Training, Outreach, Research and Education at the University of Texas, Austin, the programme has a three-year US\$994,702 grant from the US Department of Energy and will become self-sustaining after the grant expires, according to project director Hilary Olson. Topics covered will include resource assessment, site characterization, reservoir engineering, geochemical and geomechanical impact and risk assessment.

International health

The number of students enrolled in global health programmes at universities across the United States and Canada has doubled since 2006, thanks to growing interest in careers addressing health disparities and disease prevention in developing countries.

These survey findings were released on 14 September by the Consortium of Universities for Global Health, an alliance of North American universities formed to organize health projects between universities and developing nations. The survey found that numbers had increased at both the undergraduate and postgraduate level — graduate students increased from 949 in 2006 to 2,010 this year, undergraduates from 1,286 to 2,687.

Structural biologist Sarah Hymowitz first considered an industry postdoctoral fellowship while working towards her PhD at the University of California, San Francisco. She was attracted by the idea of getting more practical training than feasible during the typical postdoctoral stint — and encouraged by her existing collaborations with industry. “You knew that what you were doing might eventually help patients, rather than just saying in the last sentence of your paper, ‘This might have some value,’” she says.

Soon after earning her doctorate, Hymowitz was accepted for a three-year fellowship at her company of interest: Genentech in South San Francisco. After she completed it, Genentech immediately hired her. Hymowitz has never considered returning to academia. “I’m having way too much fun here to ever go back,” she says.

Hymowitz enjoys the collaborative nature of industry research, and appreciates the fact she has more resources — more funds, better laboratory facilities and infrastructure, the newest technologies — than she would at a university. “Here, I have to justify my projects — we all have a budget — but I have to justify them to people who want to find them interesting,” she says, referring to a strong sense of teamwork and solidarity that underpins the company. “Plus I have the fun of seeing things that I contributed to move into the clinical setting. I don’t think I could duplicate that in academia.”

Hymowitz is among a growing number of doctoral recipients in the United States who are opting to do their postdoctorate in industry. In 1981, 8% of US postdocs completed a postdoctorate at a for-profit or non-profit company. By 2005, that figure had more than tripled to 26%, according to data from the US National Science Foundation. Those who choose this path typically want to develop direct applications for their research that might have an impact on people’s lives. “If you think you want to go into industry, you’ll want to do an industrial postdoc,” says Lori Conlan, director of the office of postdoctoral services at the US National Institutes of Health.

Genentech, which has offered a postdoc programme since the company was founded more than 30 years ago, typically employs about 110 postdocs. Of those who have completed a fellowship, 10% have returned to academia. Of the 90% who have stayed in industry, 10% of those, such as Hymowitz, have been hired by Genentech.

Although many companies’ postdoc programmes focus on product development or applied science projects, Genentech’s concentrates on basic research, says Vishva Dixit, the company’s vice-president for



Industrial endeavours

Biotechnology and drug companies are piquing graduate-student interest with goal-oriented postdoctoral fellowships that maintain academic ties. **Karen Kaplan** surveys the offerings.

early discovery research and director of the postdoc programme. Genentech programme leaders believe that investing in undirected ‘blue-sky’ research is vital, and that assigning postdocs to basic research is an effective way to leverage that investment.

“I don’t consider ourselves to be emblematic of industry,” Dixit says. “No one from management can ask what a postdoc’s work has to do with the mission of the company. They are free to work on whatever intrigues them.” Management does, however, look for evidence of productivity, ideally in the form of a stellar publication record. The company benefits from its own philanthropy: such undirected research

occasionally results in the discovery of a therapeutic product, Dixit says.

Eliminating bias

Why don’t more people do industrial postdocs? Not everyone knows they exist, for one thing, nor where to look for them (see ‘Where to find industry fellowships’). Many companies that offer such programmes

advertise them only on their own websites or rely only on word of mouth.

Potentially more damning is the anti-industry feeling that exists, at least in some academic circles. Some academic administrators and faculty members may fear that postdocs who complete industrial fellowships are ill-equipped for academic research posts. Thanks to the proprietary nature of industrial research, postdocs at many companies sometimes don’t publish as often as their academic counterparts, and they typically don’t learn how to seek external funding — a fundamental component of academic research.

But this may be starting to change. Mindful that some who complete an industry postdoc may wish to return to academia, a number of biotechnology and drug companies that offer fellowships have restructured their programmes. These firms have created ways for their postdocs to maintain close links to academia, usually through academic mentors and, in at least one case, the option of doing their research at a mentor’s university.

Novartis, based in Basel, Switzerland, which employs 105 postdocs worldwide, realigned its postdoc programme in 2004 to

“I see things that I contributed to move into the clinical setting.”
— Sarah Hymowitz

more closely resemble the academic model. The company now sets up its postdocs with both an academic and in-house mentor, and requires them to design their own research proposal, which will form the basis for their postdoctoral work. About 8% of Novartis postdocs have returned to academia, according to Rajesh Ranganathan, head of the education office of Novartis Institutes for Biomedical Research. Several have been hired by the company. Ranganathan says his department is well aware of the persistent bias against industry. "We are not producing second-rate scientists who can't hack it in academia," he says.

Al Crumbliss, dean of the natural sciences at Duke University in Durham, North Carolina, says he wouldn't rule out hiring a postdoc for a faculty post simply because he had received his fellowship in industry. "The critical thing you're looking for in postdoctoral experience is increased scientific maturity, the ability to identify an important problem, and the second thing is how you're going to approach working on that problem," says Crumbliss. "There's no reason why that couldn't be developed in an industrial environment," he says. Crumbliss does, however, worry that industry postdocs won't learn how to run an academic research programme or do the difficult work of securing external funding.

Linking to academia

Industry postdoc models that include academic collaboration help address these concerns. Roche, also based in Basel, offers a dual mentorship to its postdocs; it also reconfigured its postdoctorate programme to be more similar to academia's by

requiring its postdocs to publish frequently; expecting them to collaboratively design their own research plan together with both mentors; and permitting them to conduct their research at their academic mentor's university if they choose, as a way to strengthen the academic link. "Our fellowships should offer the same scientific challenges and professional recognition as those in academia," says René Imhof, global head of scientific and talent relations, referring to the programme's collaborative research structure and regular publishing requirement. The company employs 40 postdocs globally.



"It made sense to combine research with using science to help people."
— Paul Ardayfio

Eli Lilly, based in Indianapolis, Indiana, maintains relationships with universities and faculty members through scientific conferences, e-mail, listservs and other electronic media, and informal meetings as a way to keep its postdocs linked to the academic realm. The company also encourages its postdocs to publish, although it isn't required. The company employs 31 postdocs globally and 80–90% of those who complete the programme apply for a full-time position. The offer rate is about 28%, the company says.

Paul Ardayfio, a Eli Lilly clinical research scientist, did his postdoc there and was offered a job in June. He knew he wanted to do more applied science than would be possible in academia. "I've had a long-standing interest in using my scientific background to help people and treat disease," he says. "It made a lot of sense to combine my interest in doing cutting-edge research with my desire to use that science to help people."

The academic bias against industry is fading, Ardayfio contends, as more

WHERE TO FIND INDUSTRY FELLOWSHIPS

Below is a selection of biotechnology firms and pharmaceutical companies worldwide that offer postdoctoral fellowships.

Eli Lilly ♦ www.lilly.com
Genentech ♦ www.gene.com
GlaxoSmithKline ♦ www.gsk.com
Johnson and Johnson ♦ www.jnj.com
Merck and Company ♦ www.merck.com
New England Biolabs ♦ www.neb.com
Novartis Institutes for BioMedical Research ♦ www.nibr.novartis.com
Roche ♦ www.roche.com

academic–industrial collaborations form, more companies present their work at scientific conferences and industrial postdocs continue to publish in collaboration with academic researchers. "The quality of the science that's coming out of the labs here is stellar," says Ardayfio, who received his doctorate from Harvard University. He was actually planning on doing an academic postdoc there until he visited Lilly. "The calibre of the science, the scientists themselves, the environment, the collaborative spirit, the technology — I was floored," says Ardayfio.

Anna Dzyakanchuk, a Roche postdoc in Basel, likewise cites hallmarks of the industry setting as reasons for pursuing an industry postdoc. She enjoys knowing that her work could eventually help people, and she likes working as part of a team. The Roche culture is much different from academia's, says Dzyakanchuk, who finds a far stronger focus on collaboration and less of a hypercompetitive atmosphere compared with her stint at the University of Berne in Switzerland, where she earned her doctorate in biochemistry. "I don't think I'll go back to academia," she says.

Karen Kaplan is assistant editor of Naturejobs.



Roche, based in Basel, Switzerland, offers a dual mentorship, bridging the academia–industry gap.